Microdata User Guide

Canadian Financial Capability Survey

2009

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

This package is designed to enable interested users to access and manipulate the microdata file for the Canadian Financial Capability Survey that was conducted by Statistics Canada in 2009. The survey was conducted with the cooperation and support of Human Resources and Skills Development Canada, Finance Canada and the Financial Consumer Agency of Canada. This manual provides information on the objectives, methodology and estimation procedures as well as the guidelines for releasing estimates based on the survey. Contained within this package are the questionnaire, and approximate variance tables with examples of their use.

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

Statistics Canada

Client Services
Special Surveys Division
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail: ssd@statcan.gc.ca

2.0 Background

This is the first Canadian Financial Capability Survey (CFCS). The need for this survey has been brought about by changing economic conditions, the variety and complexity of financial products available and the need to establish baseline data. The CFCS was conducted between February and May 2009 with the objective to develop a statistical database providing estimates surrounding the issues of financial capability. The intention of the survey is to collect information that will illuminate the degree of knowledge that Canadians have concerning financial decision-making. Specifically, the survey will shed light on 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 survey is designed to collect information surrounding respondents' approaches to day-to-day money management and budgeting, longer term money management and general financial planning. Information for the survey was collected from Canadians 18 years of age and older in the ten provinces.

3.0 Objectives

The fundamental objective of the Canadian Financial Capability Survey (CFCS) is to gain a greater understanding of the financial knowledge, preferences, and financial needs of Canadians. In particular the survey will collect information on Canadians' financial knowledge and understanding, their financial skills (ability to apply knowledge and make financial decisions), and their financial responsibility (behaviour in financial matters).

The information obtained from the CFCS will help governments and industry better understand the knowledge and behaviours of Canadians with respect to participation in financial service markets and in various government programs designed to facilitate financial planning for education and retirement. For example, the Registered Education Savings Program (RESP) Registered Retirement Savings Plan (RRSP) and programs such as the Guaranteed Income Supplement (GIS), and the Canada Pension Plan (CPP) / Quebec Pension Plan (QPP).

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used.

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.

Canada Pension Plan (CPP) / Quebec Pension Plan (QPP) – Retirement pensions received at age 65 by people who have worked in Canada. Also includes Survivors Benefits, such as widows' pensions, widowers' pensions, orphans' benefits and Disability Pensions for disabled pensioners.

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

Credit Union – A co-operative financial institution that is owned by its members and that operates for the benefit of its members by accepting savings deposits and making loans, including mortgage loans, and providing other services, such as chequing and credit-card services.

Co-operatives / Co-ops – see Credit Union above

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.

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).

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.

[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.

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

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.

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.

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.

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.

Net Value – Value that results from deducting operating costs from price.

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.

Occupation - The Canadian Financial Capability Survey provides information about the occupation attachment of employed and unemployed persons, and of persons not in the labour force who have held a job in the past 12 months. These codes sets are based on the National Occupational Classification – Statistics (NOC-S) 2006.

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.

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

Person Most Knowledgeable (PMK) – The PMK is the person in the household that is most knowledgeable concerning financial issues. Respondents were asked to self-identify with respect to the PMK in two sections of the questionnaire – Ongoing Expenses and Financial Management.

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.

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.

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

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.

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.

[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.

5.0 Survey Methodology

The Canadian Financial Capability Survey (CFCS) was administered between February 11th and May 9th, 2009 as a Random Digit Dialling (RDD) survey, a technique whereby telephone numbers are generated randomly by computer. Interviewing was conducted over the telephone.

5.1 Population Coverage

The target population for the CFCS was all persons 18 years of age and over living in Canada with the following two exceptions:

1) residents of the Yukon, Northwest Territories and Nunavut, and
2) full-time residents of institutions.

Because the survey was conducted using a sample of telephone numbers, households (and thus persons living in households) that do not have telephone land lines were excluded from the sample population. This means that people without telephones and people with cell phones only, were excluded. People without land lines account for about 8% of the target population. However, the survey estimates have been weighted to include persons without land lines.

5.2 Stratification

In order to ensure that people from all parts of Canada were represented in the sample, each of the 10 provinces were divided into strata or geographic areas. Census Metropolitan Areas (CMA) are areas defined by the Census of Population and correspond roughly to the cities with populations of 100,000 or more. Many CMAs were each considered as a separate stratum. This was the case for St. John's, Halifax, Saint John, Montreal, Quebec City, Toronto, Ottawa, Hamilton, Winnipeg, Regina, Saskatoon, Calgary, Edmonton, and Vancouver. The remaining CMAs in Ontario, Quebec, and British Columbia were combined into two separate strata. Generally, within each province, a non-CMA stratum was created though in Prince Edward Island there was only one stratum for the entire province. This resulted in a design with 27 strata in all.

5.3 Sample Design and Allocation

The sample design is a two-phase stratified random sample of telephone numbers. In the first phase, households are selected using RDD. In the second phase, one individual from the contacted household is selected.

Because the survey is mainly intended to produce reliable estimates at the national level, but also strives for provincial and CMA-level estimates of reasonable quality, a Kish allocation was used. As a result of this compromise, there are more respondents in the sample from the larger provinces; but the number is not strictly proportional to the population in each.

The initial sample size of telephone numbers depended upon the expected response rate and the expected RDD hit rate (proportion of sampled telephone numbers which are screened in as households). It was estimated that a total of more than 53,000 telephone numbers was needed to obtain 20,000 respondents. This assumed a 66% response rate and hit rate that varied substantially by province, with an expected overall average of about 40%.

5.4 Sample Selection

The sample for the CFCS was generated using a refinement of RDD sampling called the Elimination of Non-Working Banks (ENWB). Within each province-stratum combination, a list of working banks (area code + next five digits) was compiled from telephone company administrative files. A working bank, for the purposes of social surveys, is defined as a bank

which contains at least one working residential telephone number. Thus, all banks with only unassigned, non-working, or business telephone numbers are excluded from the survey frame.

Next, a systematic sample of banks (with replacement) was selected within each stratum. For each selected bank, a two-digit number (00 to 99) was generated at random. This random number was added to the bank to form a complete telephone number. This method allowed listed and unlisted residential numbers as well as business and non-working numbers (i.e. not currently or never in service), to have a chance of being in the sample. A screening activity aimed at removing not in service and known business numbers was performed prior to sending the sample to the computer-assisted telephone interviewing (CATI) unit.

Each telephone number in the CATI sample was dialled to determine whether or not it reached a household. If the telephone number is found to reach a household, the person answering the telephone was asked to provide information on the individual household members. The ages of the household members were used to determine who, in the household, would be selected for the interview. Respondents were interviewed in the official language of their choice and interviews by proxy respondents were not permitted.

5.5 Sample Size

The following table shows the number of households in the CFCS sample.

Province	Sample Size
Newfoundland and Labrador	1,224
Prince Edward Island	572
Nova Scotia	1,326
New Brunswick	1,213
Quebec	5,824
Ontario	8,185
Manitoba	1,427
Saskatchewan	1,976
Alberta	2,748
British Columbia	3,061
Canada	**27,555**

5.6 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. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used. 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 shell table to build the household roster. Based on the household roster an eligible household member is randomly selected to complete the survey.

Demography (DM)
This section provides some basic demographic information.

Labour force (LF)
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 (OE)
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 (FM)
This section of the questionnaire deals with longer-term financial planning. It involves questions surrounding major expenditures over ten thousand dollars, retirement planning and planning for children's post-secondary education.

Major expenses (ME)
This section of the questionnaire asks questions about how respondents are planning for future purchases or major expenditures such as a home, a new car, a cottage, or a child's upcoming wedding.

Postsecondary education funding (EF)
This section of the questionnaire asks questions about financial plans for any child or children in the event that they pursue postsecondary education such as college (collège d'enseignement général et professionnel (CEGEP) in Quebec), university or a trade apprenticeship or vocational school.

Retirement planning (RP)
This section contains questions about plans for retirement.

Assets and debts (AD)
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 (IN)
This information provides contextual information surrounding both the individual and where appropriate the household's income.

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

Subjective personal assessment (SA)
In this section of the survey, respondents provide a self-assessment of their comfort with financial matters.

Objective personal assessment (OA)
The final section of the questionnaire asks respondents to provide answers to a short money quiz.

6.0 Data Collection

Data collection for the Canadian Financial Capability Survey (CFCS) was carried out between February and early May 2009.

6.1 Questionnaire Design

In the case of the Canadian Financial Capability Survey, it was proposed from conception 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, of course, 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.

6.2 Interviewing

Statistics Canada interviewers are employees hired and trained to carry out the household surveys. The interviewers conducting the CFCS were given specific training in preparation for the survey.

Data were collected using computer-assisted telephone interviewing. A front-end module contains a set of standard response codes for dealing with all possible call outcomes, as well as the associated scripts to be read by the interviewers. A standard approach set up for introducing the agency, the name and purpose of the survey, the survey sponsors, how the survey results will be used, and the duration of the interview was used.

The CATI application ensured that only valid question responses were entered and that all the correct flows were followed. Edits were built into the application to check the consistency of responses, identify and correct outliers, and to control who gets asked specific questions. This meant that the data was already quite "clean" at the end of the collection process.

6.3 Supervision and Quality Control

All Statistics Canada 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 survey, and also for periodically monitoring their interviewers and reviewing their completed documents. The senior interviewers are, in turn, under the supervision of the program managers, located in each of the Statistics Canada regional offices.

7.0 Data Processing

The main output of the Canadian Financial Capability Survey (CFCS) 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 (computer-assisted telephone interviewing) 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 (soft-edit), 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

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”.

7.3 Coding of 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 – 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”.

7.4 Imputation

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

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

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

7.5 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. These variables and were grouped in pre-determined intervals to aid the analytical process.

7.6 Weighting

The principle behind estimation in a probability sample such as the CFCS 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 people in Canada who do not have a personal bank account is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic (OE_Q02A = 0) and summing the weights entered on those records.

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

7.7 Suppression of Confidential Information

It should be noted that the "Public Use" Microdata Files (PUMF) may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of file variables, grouping values into wider categories, and coding specific values into the "Not stated" category. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates by province for the Canadian Financial Capability Survey (CFCS).

Province	Total number of households	Total persons responding	Overall response rate
Newfoundland and Labrador	1,224	670	54.7
Prince Edward Island	572	324	56.7
Nova Scotia	1,326	779	58.8
New Brunswick	1,213	684	56.4
Quebec	5,824	3,336	57.3
Ontario	8,185	4,519	55.2
Manitoba	1,427	795	55.7
Saskatchewan	1,976	1,213	61.4
Alberta	2,748	1,690	61.5
British Columbia	3,061	1,509	49.3
Canada	**27,555**	**15,519**	**56.3**

A **respondent** has the following characteristics:

- The household roster was completed with no individual age refusals.
- The selected person was 18 years of age or older at the time of the interview (confirmed with the selected person).
- The selected person answered at least two-thirds of the key items in three out of the first five modules of the survey questionnaire (Demography, Labour force, Ongoing expenses, Financial management, and Major expenses) and at least one key item from the Financial management module.

8.2 Survey Errors

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

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

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and

processing cycle to monitor the quality of the data. These measures include cognitive testing to ensure concepts were clear, extensive training of interviewers with respect to the survey procedures and computer-assisted telephone interviewing (CATI) application, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions and testing of the CATI application to ensure that range checks, edits and question flow were all programmed correctly.

8.2.1 Data Collection

Interviewer training consisted of reading the CFCS Supervisor's Manual, Procedures Manual and Interviewer's Manual, practicing with the CFCS 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. The collection period ran from February 11 to May 9, 2009.

8.2.2 Data Processing

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.

8.2.3 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, values were imputed when these were missing.

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.

In total, almost 10,000 respondents (63%) were eligible donors having reported both household and personal incomes. Respondents who did not provide a dollar estimate of their incomes were asked questions in order to derive an income range. Almost 2,000 respondents (13%) did not provide any information on their incomes. The reported income ranges and the missing income information were imputed by the donor values in a series of steps, depending on the information available for other variables involved in forming the imputation groups. In a final step, the income values, whether reported or imputed, were converted into quartiles, quintiles, and deciles to assist in the analysis of survey results.

	Personal Income	Household Income
Imputed	4,775	5,223
Total	15,519	15,519
Rate (%)	30.8	33.7

The CFCS 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.

8.2.4 Measurement of Sampling Error

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

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

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

For example, suppose that, based upon the survey results, one estimates that 13.5% of Canadians report that they check their personal or joint account balances daily, and this estimate is found to have a standard error of 0.0034. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0034}{0.135} \right) X\ 100\ \%\ =\ 2.5\%$$

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

9.0 Guidelines for Tabulation, Analysis and Release

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

9.1 Rounding Guidelines

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

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Canadian Financial Capability Survey (CFCS) 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.

9.3 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.

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 Canadian adults who have a household budget or the proportion who 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: Do you have a household budget?
R: Yes / No

Q: As of today, what percent of the total price have you managed to save for your future home?
R: Less than 5% / 5 to 10% / 11 to 20% / 21 to 50% / 51 to 75% / 76 to 100%

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 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.

Examples 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

Q: For how many more years do you expect to make mortgage payments on your principal residence?
R: |_|_|_| years

9.3.3 Tabulation of Categorical Estimates

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

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

9.3.4 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of personal bank accounts held by adult women in Canada multiply the value reported in question OE_Q02A (number of personal chequing or savings accounts) 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 ($\hat{X}$) as described above,
b) estimate the women in Canada ($\hat{Y}$) in this category by summing the final weights of all records with SEX = 2, then
c) divide estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.4 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 = 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 = men;

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 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 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-off's

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

For example, the table shows that the quality of a weighted estimate of 15,000 people possessing a given characteristic in Newfoundland and Labrador is marginal.

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

Province	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Newfoundland and Labrador	27,500 & over	7,000 to < 27,500	under 7,000
Prince Edward Island	13,000 & over	3,500 to < 13,000	under 3,500
Nova Scotia	43,000 & over	11,000 to < 43,000	under 11,000
New Brunswick	46,000 & over	12,000 to < 46,000	under 12,000
Quebec	93,000 & over	23,000 to < 93,000	under 23,000
Ontario	113,500 & over	28,000 to < 113,500	under 28,000
Manitoba	52,500 & over	13,500 to < 52,500	under 13,500
Saskatchewan	33,000 & over	8,500 to < 33,000	under 8,500
Alberta	77,000 & over	19,500 to < 77,000	under 19,500
British Columbia	129,500 & over	32,500 to < 129,500	under 32,500
Canada	**99,000 & over**	**24,500 to < 99,000**	**under 24,500**

10.0 Approximate Sampling Variability Tables

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

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

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

Province	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.32	670	410,773
Prince Edward Island	1.21	324	109,723
Nova Scotia	1.29	779	750,481
New Brunswick	1.53	684	602,590
Quebec	1.39	3,336	6,167,275
Ontario	1.39	4,519	10,161,677
Manitoba	1.32	795	911,804
Saskatchewan	1.47	1,213	771,190
Alberta	1.33	1,690	2,740,104
British Columbia	1.54	1,509	3,586,822
Canada	**1.60**	**15,519**	**26,212,439**

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

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

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

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

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

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

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

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of 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.

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

The following examples based on the 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 (women), $\hat{X}_2$ is estimate 2 (men), 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 $\hat{d} = 0.351 - 0.389 = -0.038$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.351)(0.030)]^2 + [(0.389)(0.028)]^2} \\ &= \sqrt{(0.000111) + (0.000119)} \\ &= 0.015\end{aligned}$$

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 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:

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

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.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of women with a household budget who report that they always stay within their budget, to men with a household budget who report that they always stay within their budget is 0.83 for Nova Scotia while it is 1.18 for Alberta. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the Nova Scotia ratio ($\hat{R}_1$) and the Alberta ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Nova Scotia ratio is 19.4% and 15.0% for Alberta.

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

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

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 0.83 – 1.18 = -0.35 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.83)(0.194)]^2 + [(1.18)(0.150)]^2} \\ &= \sqrt{(0.0259) + (0.0313)} \\ &= 0.239\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.239 / (-0.35) = -0.683.

4) So the approximate coefficient of variation of the difference between the estimates is 68.3%. 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.

Canadian Financial Capability Survey, 2009

Approximate Sampling Variability Tables - Nova Scotia - 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	******	110.9	110.3	108.6	105.7	102.7	99.7	96.5	93.2	89.8	86.3	78.8	61.0	35.2
2	******	78.4	78.0	76.8	74.7	72.6	70.5	68.2	65.9	63.5	61.0	55.7	43.2	24.9
3	******	64.0	63.7	62.7	61.0	59.3	57.5	55.7	53.8	51.9	49.8	45.5	35.2	20.3
4	******	55.4	55.2	54.3	52.9	51.4	49.8	48.2	46.6	44.9	43.2	39.4	30.5	17.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
50	******	******	******	******	14.9	14.5	14.1	13.6	13.2	12.7	12.2	11.1	8.6	5.0
55	******	******	******	******	**14.3**	13.9	13.4	13.0	12.6	12.1	11.6	10.6	8.2	4.8
60	******	******	******	******	13.6	13.3	12.9	12.5	12.0	11.6	11.1	10.2	7.9	4.5
65	******	******	******	******	**13.1**	12.7	12.4	12.0	11.6	11.1	10.7	9.8	7.6	4.4
70	******	******	******	******	12.6	12.3	11.9	11.5	11.1	10.7	10.3	9.4	7.3	4.2
75	******	******	******	******	12.2	11.9	11.5	11.1	10.8	10.4	10.0	9.1	7.0	4.1
80	******	******	******	******	******	11.5	11.1	10.8	10.4	10.0	9.6	8.8	6.8	3.9
85	******	******	******	******	******	11.1	10.8	10.5	10.1	9.7	9.4	8.5	6.6	3.8
90	******	******	******	******	******	10.8	10.5	10.2	9.8	9.5	9.1	8.3	6.4	3.7
95	******	******	******	******	******	10.5	10.2	9.9	9.6	9.2	8.9	8.1	6.3	3.6
100	******	******	******	******	******	10.3	10.0	9.6	9.3	9.0	8.6	7.9	6.1	3.5
125	******	******	******	******	******	******	8.9	8.6	8.3	8.0	7.7	7.0	5.5	3.2
150	******	******	******	******	******	******	8.1	7.9	7.6	7.3	7.0	6.4	5.0	2.9
200	******	******	******	******	******	******	******	******	6.6	6.4	6.1	5.6	4.3	2.5
250	******	******	******	******	******	******	******	******	******	5.7	5.5	5.0	3.9	2.2
300	******	******	******	******	******	******	******	******	******	******	5.0	4.5	3.5	2.0
350	******	******	******	******	******	******	******	******	******	******	******	4.2	3.3	1.9
400	******	******	******	******	******	******	******	******	******	******	******	******	3.1	1.8
450	******	******	******	******	******	******	******	******	******	******	******	******	2.9	1.7
500	******	******	******	******	******	******	******	******	******	******	******	******	2.7	1.6

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

Canadian Financial Capability Survey, 2009

Approximate Sampling Variability Tables - Alberta - 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	146.7	146.1	145.3	143.1	139.3	135.3	131.3	127.1	122.8	118.4	113.7	103.8	80.4	46.4
2	103.8	103.3	102.8	101.2	98.5	95.7	92.8	89.9	86.8	83.7	80.4	73.4	56.9	32.8
3	******	84.3	83.9	82.6	80.4	78.1	75.8	73.4	70.9	68.3	65.7	59.9	46.4	26.8
4	******	73.0	72.7	71.5	69.6	67.7	65.7	63.6	61.4	59.2	56.9	51.9	40.2	23.2
5	******	65.3	65.0	64.0	62.3	60.5	58.7	56.9	54.9	52.9	50.9	46.4	36.0	20.8
6	******	59.6	59.3	58.4	56.9	55.3	53.6	51.9	50.1	48.3	46.4	42.4	32.8	19.0
7	******		55.2	54.9	54.1	52.6	51.2	49.6	48.1	46.4	44.7	43.0	39.2	30.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
100	******	******	******	14.3	13.9	13.5	13.1	12.7	12.3	11.8	11.4	10.4	8.0	4.6
125	******	******	******	12.8	12.5	12.1	11.7	11.4	11.0	10.6	10.2	9.3	7.2	4.2
150	******	******	******	******	**11.4**	11.1	10.7	10.4	10.0	9.7	9.3	8.5	6.6	3.8
200	******	******	******	******	**9.8**	9.6	9.3	9.0	8.7	8.4	8.0	7.3	5.7	3.3
250	******	******	******	******	8.8	8.6	8.3	8.0	7.8	7.5	7.2	6.6	5.1	2.9
300	******	******	******	******	******	7.8	7.6	7.3	7.1	6.8	6.6	6.0	4.6	2.7
350	******	******	******	******	******	7.2	7.0	6.8	6.6	6.3	6.1	5.5	4.3	2.5
400	******	******	******	******	******	6.8	6.6	6.4	6.1	5.9	5.7	5.2	4.0	2.3
450	******	******	******	******	******	******	6.2	6.0	5.8	5.6	5.4	4.9	3.8	2.2
500	******	******	******	******	******	******	5.9	5.7	5.5	5.3	5.1	4.6	3.6	2.1
750	******	******	******	******	******	******	******	******	4.5	4.3	4.2	3.8	2.9	1.7
1,000	******	******	******	******	******	******	******	******	******	******	3.6	3.3	2.5	1.5
1,500	******	******	******	******	******	******	******	******	******	******	******	******	2.1	1.2
2,000	******	******	******	******	******	******	******	******	******	******	******	******	******	1.0

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

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

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

> If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.
>
> Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the difference would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.
>
> Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X} - k, \hat{X} + k\right)$ where k is

determined depending upon the level of confidence desired and the sampling error of the estimate.

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

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

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

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

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

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

A 95% confidence interval for the estimated proportion of 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.

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

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers,

averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

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

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

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

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of 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.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the CFCS 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 personal bank accounts held by women would be greater than the coefficient of variation of the corresponding proportion of women with a personal bank account. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

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

10.5 Coefficient of Variation Tables

Refer to CFCS2009_CVTabsE.pdf for the coefficient of variation tables.

10.6 Mean Bootstrap Method for Variance Estimation

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 mean bootstrap method is used because the sample design and calibration needs to be taken into account when calculating variance estimates. The mean bootstrap method does this, and with the use of the Bootvar program, discussed in the next section, is a method that is fairly easy for users.

The CFCS uses the mean bootstrap method described by W. Yung (Yung, W. (1997b). Variance estimation for public use microdata files. *Proceedings of Symposium 1997: New Directions in Surveys and Censuses*, Statistics Canada).

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. The CFCS uses $R = 20$ and $B = 250$, to produce 250 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 250 final mean bootstrap weights for each unit in the sample. The variation among the 250 possible estimates based on the 250 mean bootstrap weights are related to the variance of the estimator based on the regular weights and can be used to estimate it. There are a number of reasons why a user may need to calculate the coefficient of variation of estimates with the mean bootstrap method. A few are given below.

- First, if a user wishes to have estimates at a geographic level smaller than the province (for example, at the urban or rural level), then the Approximate Sampling Variability Tables provided are not adequate. Coefficients of variation of these estimates may be obtained using "domain" estimation techniques through the Bootstrap variance program.

- Second, should a user require more sophisticated analyses such as estimates of coefficients from linear regressions or logistic regressions, the Approximate Sampling Variability Tables will not provide correct associated coefficients of variation. Although some standard statistical packages allow sampling weights to be incorporated in the analyses, the variances that are produced often do not properly take into account the design and/or calibration of the weights, whereas the Bootstrap variance program does.

- Third, for estimates of quantitative variables, separate tables are required to determine their sampling error.

10.7 Statistical Packages for Variance Estimation

Statistics Canada has developed a program that can perform bootstrap variance estimation: the Bootvar program.

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

11.0 Weighting

For the microdata file, statistical weights were placed on each record to represent the number of sampled persons that the record represents. One weight was calculated for each responding person.

The weighting for the Canadian Financial Capability Survey (CFCS) consisted of several steps:

- calculation of a basic weight,
- adjustments for non-response,
- dropping out-of-scope records,
- an adjustment for selecting one individual in the household, and finally,
- an adjustment to make the populations estimates consistent with known province-age-sex totals from the Census projected population counts for persons 18 years and over.

11.1 Weighting Procedures

1. Calculate telephone weight

Each telephone number in the sample was assigned a basic weight, W_1, equal to the inverse of its probability of selection.

$$W_1 = \left(\frac{Total\ number\ of\ possible\ sampled\ telephone\ numbers\ in\ province}{Number\ of\ sampled\ telephone\ numbers\ in\ province} \right)$$

There were 68,462 telephone numbers in the sample with assigned weights.

2. Adjust for non-resolved telephone numbers

There were 5,191 telephone numbers that were not resolved, leaving 63,271 resolved telephone numbers. The unresolved telephone numbers were not determined to belong to a household, business or out-of-scope. Each telephone number had a flag indicating whether it was expected to be a residential, business, or unknown type of telephone number, and a flag indicating whether or not it was screened out before collection as a non-working or business number. The adjustment for the unresolved telephone numbers was done within province, the expected line type, and whether or not the number was sent to the field.

For each province expected line type sent,

$$W_2 = W_1 * \left(\frac{\sum W_1\ for\ resolved\ telephone\ numbers + \sum W_1\ for\ unresolved\ telephone\ numbers}{\sum W_1\ for\ resolved\ telephone\ numbers} \right)$$

3. Remove out-of-scope telephone numbers

Telephone numbers corresponding to businesses, out-of-service numbers, or out-of-scope numbers, such as cottage telephone numbers, were dropped after the non-resolved adjustment had been applied. Note that if household or person data existed then the telephone number was assumed to be a household. There were 38,040 out-of-scope telephone numbers and 25,231 telephone numbers belonging to a household.

4. Adjust for non-response of number of telephone lines in the household

The number of telephone lines in the household was calculated. If the number of different telephone lines within the household could not be calculated but household or person data existed, then it was imputed as one in order to retain good data. After imputation, there were 7,384 telephone numbers that were still missing the number of lines. Thus, there were 17,847 households with the number of lines calculated or imputed. The adjustment was done within province.

$$W_3 = W_2 * \left(\frac{\sum W_2 \ for\ households\ with\ number\ of\ lines + \sum W_2 \ for\ households\ missing\ number\ of\ lines}{\sum W_2 \ for\ households\ with\ number\ of\ lines} \right)$$

5. Calculate household weight with multiple telephone lines adjustment

Weights for households with more than one telephone line (with different telephone numbers) were adjusted downwards to account for the fact that such households have a higher probability of being selected. The weight for each household was divided by the number of distinct residential telephone lines (up to a maximum of 4) that serviced the household. The adjustment was done within province.

$$W_4 = \left(\frac{W_3}{Number\ of\ in-scope\ telephone\ lines\ in\ the\ household} \right)$$

6. Adjust for non-responding households

Household respondents responded to the questions used to create the household roster. If these questions were not sufficiently answered, perhaps refused or only partially answered, then the household was considered a non-respondent. There were 71 non-respondents. Thus, 17,776 in-scope household weights were used and adjusted within province.

$$W_5 = W_4 * \left(\frac{\sum W_4 \ for\ household\ respondents + \sum W_4 \ for\ household\ non-respondents}{\sum W_4 \ for\ household\ respondents} \right)$$

7. Assign household weights to selected persons

All selected persons from the in-scope responding households with completed rosters (i.e. no missing ages) were assigned their household weights.

$$W_6 = W_5$$

8. Calculate selected person sub-weight

The weight for each selected person is then inflated using the roster information to represent the number of people within the household who were eligible to be selected (aged 18 years or older).

$$W_7 = W_6 * Number\ of\ eligible\ household\ members$$

9. Adjust for non-responding individuals

The data file includes records of individual respondents who completed a sufficient amount of the questions asked. There were 2,257 non-respondents.

Thus, 15,519 in-scope individual weights were used and adjusted within province, age groups derived from the roster (18 to 24, 25 to 44, 45 to 64, 65 and over) and sex.

$$W_8 = W_7 * \left(\frac{\sum W_7 \ for \ person \ respondents + \sum W_7 \ for \ person \ non-respondents}{\sum W_7 \ for \ person \ respondents} \right)$$

10. Adjust to external totals

An adjustment was made to the person weights in order to make population estimates consistent with external population counts for persons 18 years and older. This is known as post-stratification. The following external control totals, as projected for February 2009, were used:

1) Population totals by province, sex and the following age groups: 18 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69 and 70 and over.

2) Population totals of persons aged 18 years and older in Census Metropolitan Areas.

The method called generalized regression (GREG) estimation was used to modify the weights to ensure that the survey estimates agreed with the external totals simultaneously along the two dimensions.

The person weights obtained after this step were considered final and appear on the microdata file.

12.0 Questionnaires

The Canadian Financial Capability Survey (CFCS) questionnaire was used in 2009 to collect information for the survey. The file CFCS2009_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See CFCS2009_CdBk.pdf for the record layout with univariate frequencies.