Microdata User Guide

The Survey of Older Workers

2008

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Survey of Older Workers (SOW) was conducted by Statistics Canada in 2008 with the cooperation and support of Human Resources and Skills Development Canada. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

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

Human Resources and Skills Development Canada

Joni Baran
140 Promenade du Portage
Gatineau, Quebec
Telephone: 819-997-5322
Fax: 819-953-8479
E-mail: joni.baran@hrsdc-rhdcc.gc.ca

2.0 Background

These data support labour policy initiatives in the federal government. This informs policy development and may assist in future decision-making.

Survey data will be used to:

- Provide national and provincial level estimates of households with workers aged 50 years and older and estimates on whether these workers experienced unemployment after the age of 50;
- Provide national level information on reasons for a spell of unemployment and what may follow a spell of unemployment;
- Provide national level estimates of the number of years working for their last employer;
- Provide national level estimates of the types of considerations that might encourage recent retirees to re-enter the labour force;
- Provide national level estimates on the main source of information in learning about retirement planning;
- Provide national level estimates of work/retirement plans by industry, and occupation;
- Provide national level estimates of work/retirement plans by type of pension plan;
- Provide national level estimates of identified barriers to saving for retirement.

3.0 Objectives

The intention of the survey is to collect information that will shed light on the intentions, plans and related factors that can affect the decisions of older workers to retire, continue working in their "career job" or to seek work in alternate occupations.

The main objectives of the survey are to identify:

- Health related factors that impact the decision to retire from a "career job", remain working, or start a new job;
- The personal financial considerations that impact on the decision to retire from a "career job", remain working or start a new job;
- The role of pension plans in the retirement/work decision;
- The role of dependents in the retirement/work decision;
- The role of occupation, industry, gender, marital status and geography in the retirement/work decision.

4.0 Concepts and Definitions

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

4.1 Labour Force Survey Concepts and Definitions

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

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

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

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

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

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

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

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

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

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

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

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

4.2 Survey of Older Workers Concepts and Definitions

Account
A financial record of receipts and disbursements, income and expenditures, credits and debits.

Annuity
A financial management approach to providing regular payments from a principle (lump sum) while still accumulating interest on the principle.

Canada Pension Plan (CPP) and Quebec Pension Plan (QPP)
These plans are contributory earnings related social insurance programs that ensure a measure of income protection to contributors and their families against loss of income due to retirement, disability or death. The CPP operates in all provinces and territories except Quebec, where the similar QPP is in effect. The CPP can be collected as early as age 60.

Career Jobs
Jobs held for 8 years or longer on a full-time basis.

Class of Worker
Indicates whether a person is an *employee* (that is, working for a wage or salary, a commission, or payment in kind); *self-employed* on a farm, in a business, or in a professional practice (either incorporated or unincorporated, with or without paid help).

Employment Insurance (EI)
Employment Insurance provides temporary financial assistance for unemployed Canadians while they look for work or upgrade their skills. Canadians who are sick, pregnant or caring for a newborn or adopted child, as well as those who must care for a family member who is seriously ill with a significant risk of death, may also be assisted by Employment Insurance.

Equity
The worth or value of the assets of a business or person after all factors have been taken into consideration, subtracted, or settled.

Full-time Workers
Are those who reported working 30 hours or more per week during most of the weeks that they were employed.

Household
Any person or group of persons living in a private dwelling. It may consist of one person living alone, or a group of people who are not related but share the same dwelling, or it may be a family.

Layoff
The act of suspending or dismissing a worker through no fault of the worker.

Occupational Pension Plan
A pension plan generated by a company, or organization for the benefit of its employees. In "contributory" plans both the employer and employee contribute to a fund which grows free of tax during the savings period. In "non-contributory" plans, only the employer contributes.

Old Age Security (OAS) and Guaranteed Income Supplement (GIS)
The OAS is a taxable monthly payment to people age 65 and older based on years of residency in Canada. The GIS is a non-taxable benefit paid to lower-income OAS recipients. Both benefits are income-tested and can be clawed back as income increases.

Older Worker
For the purposes of the Survey of Older Workers, an individual must be between 50 and 75 years of age and active in the labour market or have been retired 24 months or less.

Other Private Savings and Investments
Any savings (in an account or otherwise) or investments (e.g., stocks, bonds and mutual funds) that a respondent has and will use as income.

Part-time Worker
Are workers who worked less than 30 hours per week. (In the case of people who had more than one job during the same week, the hours spent at all jobs were combined).

Registered Retirement Income Fund (RRIF)
A RRIF is for individuals, established at a financial institution, and registered under the Income Tax Act to provide streamed income in retirement while still collecting interest on the principle. They typically function like an annuity. RRIFs are set up by directly transferring monies from Registered Retirement Savings Plans or lump-sum payments from Registered Pension Plans. Amounts withdrawn from RRIFs are taxable. A minimum amount must be withdrawn each year, beginning in the year after the RRIF is established.

Registered Pension Plan (RPP)
A RPP is sponsored by an employer or union and is usually funded through both employee and employer contributions. RPPs must satisfy certain conditions and be registered for the purposes of the federal *Income Tax Act*. Contributions to RPPs are tax deductible, the investment income in them is tax deferred, and payments from them are taxable.

Registered Retirement Savings Plan (RRSP)
RRSPs provide retirement income based on previously accumulated contributions and interest accrued on those contributions. RRSPs are purchased either by the individual or spouse. Employment status is not a factor that is relevant for the purchase of RRSPs. Contributions to an RRSP are tax-deductible, and the investment income is tax-deferred. RRSPs can be cashed in at anytime and are taxed accordingly. You can make contributions up to and including your 71st year. You must convert an RRSP to a RRIF (registered retirement income fund) before the end of the year in which you become 71 years of age.

Retired
In the context of this survey, refers to a person who is aged 50 and older, has left their main job or employer, and receives 50% or more of his or her total income from retirement-like sources. It is possible to retire and then return to the workforce in another capacity.

Reverse Mortgage
A reverse mortgage allows the homeowner to access equity in their home while still living in the home.

Superannuation
A pension or payment to a person retiring from full-time work on reaching the age of retirement. The term also refers to the accumulated contributions by employers and employees to a superannuation fund.

Veteran's Pension
The *Pension Act* provides pension awards to those suffering from disabilities related to military service, either during peace or war time. The pension award is based on the extent of the disability, as verified by a medical examination, and paid in accordance with rates set out in the Act.

When a disability pensioner dies, the spouse/common-law partner may receive, for a period of one year, the same pension amount being paid to the pensioner at the time of death. After the one year, a survivor will receive a survivor's pension. Surviving children may be eligible for orphan benefits following a pensioner's death. Surviving spouses/surviving common-law partners who remarry will continue to receive survivor benefits.

Workers Compensation
Workers' Compensation programs protect employees from the financial hardships associated with work-related injuries and occupational diseases.

Workplace Pension
A pension plan generated by a company, or organization for the benefit of its employees. In 'contributory' plans both the employer and employee contribute to a fund which grows free of tax during the savings period. In 'non-contributory' plans, only the employer contributes. While there are a variety of workplace pensions they are typically defined contribution or defined benefit.

5.0 Survey Methodology

The Survey of Older Workers (SOW) was administered in October and November 2008 to a sub-sample of the dwellings in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. The LFS design is briefly described in the Sections 5.1 to 5.4.[3] Sections 5.5 and 5.6 describe how the SOW departed from the basic LFS design in fall 2008.

5.1 Population Coverage

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

5.2 Sample Design

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

5.2.1 Primary Stratification

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

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

5.2.2 Types of Areas

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

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

[3] A detailed description of the LFS design is available in the Statistics Canada publication entitled *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

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

5.2.3 Secondary Stratification

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

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

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

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

5.2.4 Cluster Delineation and Selection

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

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

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

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

5.2.5 Dwelling Selection

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

5.2.6 Person Selection

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

5.3 Sample Size

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

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

5.4 Sample Rotation

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

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

5.5 Modifications to the Labour Force Survey Design for the Survey of Older Workers

The SOW used five of the six rotation groups in the LFS sample. Rotation groups 2, 5 and 6 were selected in October and 1 and 3 in November. For the SOW, the coverage of the LFS was modified to include all members of the household aged 50 to 75 who were either still working or who had retired within the last 24 months. However, unlike the LFS where information is

collected for all eligible household members, the SOW only collected information from one pre-selected household member and proxy responses were not permitted.

5.6 Sample Size by Province for the Survey of Older Workers

The following table shows the number of respondents in the LFS sampled rotations who were eligible for the SOW supplement.

Province	Sample Size
Newfoundland and Labrador	484
Prince Edward Island	347
Nova Scotia	707
New Brunswick	635
Quebec	2,263
Ontario	3,641
Manitoba	933
Saskatchewan	927
Alberta	1,192
British Columbia	1,489
Canada	**12,618**

6.0 Data Collection

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

6.1 Interviewing for the Labour Force Survey

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

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

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

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

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

6.2 Supervision and Quality Control

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

6.3 Non-response to the Labour Force Survey

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

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

6.4 Data Collection Modifications for the Survey of Older Workers

The Survey of Older Workers (SOW) was administered to one randomly selected individual per household. The random selection was carried out at the time of the interview.

Upon completion of the Labour Force Survey interview, the interviewer asked to speak to the selected person for the SOW. If the selected person was not available, the interviewer arranged for a convenient time to phone back. Proxy response was not allowed unless it was to aid in communicating with the selected respondent (i.e., translation, disability, etc.), and the selected respondent was present. Hence the collection period was extended by one week to allow the interviewers time to contact the pre-selected individuals.

6.5 Non-response to the Survey of Older Workers

For households responding to the LFS, the next stage of data collection was to administer the SOW. The interviews for the SOW were done, when possible, at the same time as the LFS interview. In some cases, the LFS and SOW respondent were different household members. If the selected respondent couldn't complete the SOW interview right away, the interviewer would make an appointment. The SOW collection continued for four additional days after the LFS interviews were completed. The SOW was always initiated during the LFS survey week. As with all supplements, the LFS interview took priority over the SOW interview.

7.0 Data Processing

The main output of the Survey of Older Workers (SOW) is a “clean” microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

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

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

7.2 Editing

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.

There are a few circumstances where respondents may be found out-of-scope of the SOW. By far, the majority of out-of-scope sampled cases are found among respondents who either did not meet the age selection criteria (aged 50 to 75 years) or the labour force participation criteria (having worked within the past 24 months prior to interview).

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 which 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. A total of 11 partially or completely open-ended questions were included in the survey. These were items relating to what could be done to entice the unemployed back to work, factors that affected

retirement or a return to work after retirement, sources of income, reasons for job displacement, etc.

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 SOW, no donor imputation was used to fill in missing data for item and partial non-response.

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 relating to age and the number of years of service. These variables were grouped into pre-determined intervals in an effort to aid the analytical process.

7.6 Weighting

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

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example if the number of individuals aged 65 to 75 who were still working is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

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

7.7 Suppression of Confidential Information

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

The survey master file includes geographic identifiers that are more explicit than the PUMF. Notably included are the economic region, the census metropolitan area, as well as the province. The PUMF does not contain any geographic identifiers below the provincial level.

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

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

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Survey of Older Workers (SOW).

Province	In-scope Sample	Responses	Response Rate (%)
Newfoundland and Labrador	484	380	78.5
Prince Edward Island	347	274	79.0
Nova Scotia	707	562	79.5
New Brunswick	635	490	77.2
Quebec	2,263	1,827	80.7
Ontario	3,641	2,947	80.9
Manitoba	933	776	83.2
Saskatchewan	927	776	83.7
Alberta	1,192	915	76.8
British Columbia	1,489	1,087	73.0
Canada	**12,618**	**10,034**	**79.5**

Note: The SOW response rate is the number of SOW responding individuals as a percentage of the number of SOW selected in-scope individuals.

8.2 Survey Errors

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

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

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort were taken to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

8.2.1 The Frame

Because the SOW was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the SOW frame. The quality of the sampling variables in the frame was very high. The SOW sample consisted of 5 rotation groups from the LFS.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the SOW frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data. The SOW frame also excludes full non-response to the LFS and item non-response to variables used in the selection criteria.

The variables on the SOW frame were quite up-to-date since they were collected from the LFS at most four weeks before the beginning of the SOW collection.

8.2.2 Data Collection

Interviewer training consisted of reading the SOW Supervisor's Manual, Procedures Manual and Interviewer's Manual, practicing with the SOW training cases on the computer for computer-assisted interviewing (CAI) only, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers collected the SOW information after the LFS information was collected. The collection period ran from the week of October 19th to November 23rd, 2008.

When possible, the interviews for the SOW were done at the same time as the LFS interview. The LFS and SOW respondent may have been different household members. If the selected respondent did not complete the SOW interview right away, the interviewer would make an appointment. The SOW would continue for four more days after the LFS survey week. The SOW was always initiated during the LFS survey week. As always for supplements, the LFS interview took priority over the SOW interview.

In an instance where the LFS is completed and the SOW is not, the SOW split into a separate group Z to be finalized.

8.2.3 Data Processing

During processing of the data, 641 SOW records did not match to corresponding records in the LFS. Thus they were coded as non-response. These records include respondents who answered the LFS but did not respond to the SOW or began the SOW but did not complete it to the point where it could be considered a useable record.

Data processing of the SOW 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.

Due to the large amount of editing built in to the computer-assisted telephone interviewing (CATI) application, the SOW dataset only required two edits.

- Firstly, 15 records had question PP_Q05 set to "Not stated" after it was found that the age at which they planned to apply for their Canada Pension Plan/Quebec Pension Plan benefits was less than their actual age.

- Second, question DW_Q03 asked respondents if they were still without work after their displacement. Question DW_Q03 was set to "Not stated" in 67 instances where the response was inconsistent with the respondent's current employment situation (CS_Q01).

No other data items were imputed, deterministically or otherwise.

Finally, in order to prevent the disclosure of a SOW respondent, all numerical variables and some coded variables have been grouped.

8.2.4 Non-response

Province	Number of non-respondents	Number of respondents	Response rate (%)	Number of known out-of-scope	Total sample size (in-scope)
Newfoundland and Labrador	104	380	78.5	0	484
Prince Edward Island	73	274	79.0	0	347
Nova Scotia	145	562	79.5	0	707
New Brunswick	145	490	77.2	0	635
Quebec	436	1,827	80.7	2	2,263
Ontario	694	2,947	80.9	7	3,641
Manitoba	157	776	83.2	0	933
Saskatchewan	151	776	83.7	1	927
Alberta	277	915	76.8	1	1,193
British Columbia	402	1,087	73.0	0	1,489
Canada	**2,584**	**10,034**	**79.5**	**11**	**12,618**

Based on a review of the findings it was evident that there were very few categories where response rates were affected by the socio-demographics of the chosen respondent. Geographically, British Columbia had a lower response rate than any other province, but this phenomenon is a current trend across the Department and is not unique to the SOW. Implementing the SOW as a voluntary supplement to the LFS resulted in a highly satisfactory response rate of 79.5%.

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.

8.2.5 Measurement of Sampling Error

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some

indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

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

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

For example, suppose that, based upon the 2008 SOW results, one estimates that 71.16% of individuals in the SOW population in Prince Edward Island are still working and have never previously retired and this estimate is found to have a standard error of 0.0311. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0311}{0.7116} \right) X \ 100 \ \% \ = \ 4.37 \ \%$$

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 Survey of Older Workers (SOW) 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 SOW 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 SOW.

9.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of persons working who are self-employed during the reference week or the proportion of persons who are retired 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: Are you currently collecting a retirement pension from Canada Pension Plan/Quebec Pension Plan?
R: Yes / No

Q: What contributed to your decision to go back to work after you retired?
R: Financial need / Family related reasons / Always planned to return to work doing something different / Social interaction, something to do / Other - Specify

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 months without work following a job loss. The numerator is an estimate of the total number of months without work and its denominator is the number of workers who have been displaced but have since found other employment.

Example of Quantitative Questions:

Q: How many years of pensionable service do you have with your current employer?
R: |_|_| years

Q: How long have you been without work?
R: |_|_|_| months

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 months without work following a displacement for persons living in Ontario, multiply the value reported in question DW_Q04 (number of months without work (000 to 168)) for all persons responding to this question by the final weight for the record, then sum this value over all records with PROV = 35 (Ontario) and CS_Q04 = 1 (lost a job following a layoff, plant or business closure, or downsizing).

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 months without work following a displacement for persons living in Ontario,

a) estimate the total number of months without work ($\hat{X}$) as described above,
b) estimate the number of persons currently working in Ontario who have been displaced ($\hat{Y}$) in this category by summing the final weights of all records with PROV = 35 and CS_Q04 = 1, then
c) divide estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.4 Guidelines for Statistical Analysis

The SOW 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. A very good way to approximate the true variance is to use a replication method, namely the bootstrap method. This method is known to correctly approximate the true value of the variance. A file containing 1,000 bootstrap weights is available. Variance calculation using 1,000 bootstrap weights involves calculating the estimates with each of these 1,000 weights and then, calculating the variance of these 1,000 estimates.

Statistics Canada has developed a statistical product called BOOTVAR which estimates variances using the bootstrap method. BOOTVAR allows for the estimation of variances for totals, ratios (including proportions), differences between ratios (or proportions), linear regression models, and logistic regression models. Version 3.1 of BOOTVAR (available in SAS only) also allows for the estimation of percentiles and Chi-square tests of independence. BOOTVAR is available in both the SAS and SPSS programming languages. The most recent version of the program (SAS version 3.1 and SPSS version 3.0), user documentation, and the details of specific survey parameters can be downloaded free of charge by following these links:

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

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the SOW, 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 for the Survey of Older Workers

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 5,000 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	9,800 & over	2,600 to < 9,800	under 2,600
Prince Edward Island	4,000 & over	1,100 to < 4,000	under 1,100
Nova Scotia	12,400 & over	3,200 to < 12,400	under 3,200
New Brunswick	12,000 & over	3,200 to < 12,000	under 3,200
Quebec	46,800 & over	11,800 to < 46,800	under 11,800
Ontario	54,300 & over	13,600 to < 54,300	under 13,600
Manitoba	13,300 & over	3,400 to < 13,300	under 3,400
Saskatchewan	11,500 & over	3,000 to < 11,500	under 3,000
Alberta	42,800 & over	11,100 to < 42,800	under 11,100
British Columbia	42,300 & over	10,800 to < 42,300	under 10,800
Canada	**44,600 & over**	**11,000 to < 44,600**	**under 11,000**

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 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 Survey of Older Workers (SOW).

Province	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.39	380	83,283
Prince Edward Island	1.45	274	25,089
Nova Scotia	1.31	562	157,387
New Brunswick	1.42	490	123,909
Quebec	2.00	1,827	1,212,979
Ontario	2.27	2,947	1,977,841
Manitoba	1.58	776	191,659
Saskatchewan	1.54	776	170,076
Alberta	2.04	915	565,817
British Columbia	1.77	1,087	750,332
Canada	**2.34**	**10,034**	**5,258,373**

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 releasing 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 retired persons receiving Canada Pension Plan/Quebec Pension Plan benefits is more reliable than the estimated number of retired persons receiving Canada Pension Plan/Quebec Pension Plan benefits. (Note that in the tables the coefficients of variation decline in value reading from left to right).

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

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

Rule 3: Estimates of Differences Between Aggregates or Percentages

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

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

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

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of persons working and the numerator is the number of persons working but previously retired.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of persons working in the public sector as compared to the number of persons working in the private sector 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 2008 SOW Master file 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 383,042 individuals in the SOW population belong to the 65 to 69 year old age group. 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 (383,042) 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 400,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 5.2%.

4) So the approximate coefficient of variation of the estimate is 5.2%. The finding that there were 383,042 (to be rounded according to the rounding guidelines in Section 9.1) persons aged 65 to 69 is publishable with no qualifications.

Survey of Older Workers, 2008

Approximate Sampling Variability Tables - Canada

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	110.6	110.1	109.5	107.8	105.0	102.0	99.0	95.8	92.6	89.2	85.7	78.2	60.6	35.0
2	78.2	77.8	77.4	76.2	74.2	72.1	70.0	67.7	65.5	63.1	60.6	55.3	42.8	24.7
3	63.8	63.6	63.2	62.3	60.6	58.9	57.1	55.3	53.4	51.5	49.5	45.2	35.0	20.2
4	55.3	55.0	54.8	53.9	52.5	51.0	49.5	47.9	46.3	44.6	42.8	39.1	30.3	17.5
5	49.5	49.2	49.0	48.2	46.9	45.6	44.3	42.8	41.4	39.9	38.3	35.0	27.1	15.6
6	******	44.9	44.7	44.0	42.8	41.6	40.4	39.1	37.8	36.4	35.0	31.9	24.7	14.3
7	******	41.6	41.4	40.8	39.7	38.6	37.4	36.2	35.0	33.7	32.4	29.6	22.9	13.2
8	******	38.9	38.7	38.1	37.1	36.1	35.0	33.9	32.7	31.5	30.3	27.7	21.4	12.4
9	******	36.7	36.5	35.9	35.0	34.0	33.0	31.9	30.9	29.7	28.6	26.1	20.2	11.7
10	******	34.8	34.6	34.1	33.2	32.3	31.3	30.3	29.3	28.2	27.1	24.7	19.2	11.1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
250	******	******	******	6.8	6.6	6.5	6.3	6.1	5.9	5.6	5.4	4.9	3.8	2.2
300	******	******	******	******	6.1	5.9	5.7	5.5	5.3	5.1	4.9	4.5	3.5	2.0
350	******	******	******	******	5.6	5.5	5.3	5.1	4.9	4.8	4.6	4.2	3.2	1.9
400	******	******	******	******	**5.2**	5.1	4.9	4.8	4.6	4.5	4.3	3.9	3.0	1.7
450	******	******	******	******	4.9	4.8	4.7	4.5	4.4	4.2	4.0	3.7	2.9	1.6
500	******	******	******	******	4.7	4.6	4.4	4.3	4.1	4.0	3.8	3.5	2.7	1.6
750	******	******	******	******	******	3.7	3.6	3.5	3.4	3.3	3.1	2.9	2.2	1.3
1,000	******	******	******	******	******	******	3.1	3.0	2.9	2.8	2.7	2.5	1.9	1.1
1,500	******	******	******	******	******	******	******	******	2.4	2.3	2.2	2.0	1.6	0.9
2,000	******	******	******	******	******	******	******	******	******	******	1.9	1.7	1.4	0.8
3,000	******	******	******	******	******	******	******	******	******	******	******	******	1.1	0.6
4,000	******	******	******	******	******	******	******	******	******	******	******	******	******	0.6

NOTE: For correct usage of these tables, please refer to the microdata documentation.

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

Suppose that the user estimates that 507,770 / 617,067 = 82.3% of 55 to 59 year olds in Ontario are working, never retired. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., 55 to 59 year olds in Ontario), it is necessary to use both the percentage (82.3%) and the numerator portion of the percentage (507,770) in determining the coefficient of variation.

3) The numerator, 507,770, does not appear in the left-hand column (the “Numerator of Percentage” column) so it is necessary to use the figure closest to it, namely 500,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 90.0%.

4) The figure at the intersection of the row and column used, namely 1.7% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 1.7%. The finding that 82.3% of 55 to 59 year olds in Ontario are working, never retired can be published with no qualifications.

Survey of Older Workers, 2008

Approximate Sampling Variability Tables - Ontario

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	123.3	122.7	122.1	120.2	117.0	113.7	110.3	106.8	103.2	99.4	95.5	87.2	67.6	39.0
2	*****	86.8	86.3	85.0	82.7	80.4	78.0	75.5	73.0	70.3	67.6	61.7	47.8	27.6
3	*****	70.9	70.5	69.4	67.6	65.7	63.7	61.7	59.6	57.4	55.2	50.4	39.0	22.5
4	*****	61.4	61.0	60.1	58.5	56.9	55.2	53.4	51.6	49.7	47.8	43.6	33.8	19.5
5	*****	54.9	54.6	53.8	52.3	50.9	49.3	47.8	46.1	44.5	42.7	39.0	30.2	17.4
6	*****	50.1	49.8	49.1	47.8	46.4	45.0	43.6	42.1	40.6	39.0	35.6	27.6	15.9
7	*****	46.4	46.1	45.4	44.2	43.0	41.7	40.4	39.0	37.6	36.1	33.0	25.5	14.7
8	*****	43.4	43.2	42.5	41.4	40.2	39.0	37.8	36.5	35.2	33.8	30.8	23.9	13.8
9	*****	40.9	40.7	40.1	39.0	37.9	36.8	35.6	34.4	33.1	31.8	29.1	22.5	13.0
10	*****	38.8	38.6	38.0	37.0	36.0	34.9	33.8	32.6	31.4	30.2	27.6	21.4	12.3
11	*****	37.0	36.8	36.2	35.3	34.3	33.3	32.2	31.1	30.0	28.8	26.3	20.4	11.8
12	*****	35.4	35.2	34.7	33.8	32.8	31.8	30.8	29.8	28.7	27.6	25.2	19.5	11.3
13	*****	34.0	33.9	33.3	32.5	31.5	30.6	29.6	28.6	27.6	26.5	24.2	18.7	10.8
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
125	*****	*****	*****	*****	10.5	10.2	9.9	9.6	9.2	8.9	8.5	7.8	6.0	3.5
150	*****	*****	*****	*****	9.6	9.3	9.0	8.7	8.4	8.1	7.8	7.1	5.5	3.2
200	*****	*****	*****	*****	*****	8.0	7.8	7.6	7.3	7.0	6.8	6.2	4.8	2.8
250	*****	*****	*****	*****	*****	7.2	7.0	6.8	6.5	6.3	6.0	5.5	4.3	2.5
300	*****	*****	*****	*****	*****	*****	6.4	6.2	6.0	5.7	**5.5**	5.0	3.9	2.3
350	*****	*****	*****	*****	*****	*****	5.9	5.7	5.5	5.3	5.1	4.7	3.6	2.1
400	*****	*****	*****	*****	*****	*****	*****	5.3	5.2	5.0	4.8	4.4	3.4	2.0
450	*****	*****	*****	*****	*****	*****	*****	**5.0**	4.9	4.7	4.5	4.1	3.2	1.8
500	*****	*****	*****	*****	*****	*****	*****	*****	4.6	4.4	4.3	3.9	3.0	**1.7**
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.5	3.2	2.5	1.4
1,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.1	1.2
1,500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.0

NOTE: For correct usage of these tables, please refer to the microdata documentation.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 236,869 / 594,791 = 39.8% of two person households in Quebec had a household income of $50,000 to $100,000, while 319,069 / 788,592 = 40.5% of two person households in Ontario had income in the same range. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the QUEBEC and ONTARIO coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for Quebec as 5.6%, and the CV of the estimate for Ontario as 5.5%.

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

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

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

That is, the standard error of the difference $\hat{d} = 0.398 - 0.405 = -0.007$ is:

$$\sigma_{\hat{d}} = \sqrt{[(0.398)(0.056)]^2 + [(0.405)(0.055)]^2}$$
$$= \sqrt{(0.0004967) + (0.0004961)}$$
$$= 0.032$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.032 / 0.007 = 4.571$

4) So the approximate coefficient of variation of the difference between the estimates is 457.1%. 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.

Survey of Older Workers, 2008

Approximate Sampling Variability Tables - Quebec

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	115.1	114.6	114.0	112.2	109.2	106.2	103.0	99.7	96.3	92.8	89.2	81.4	63.1	36.4
2	*****	81.0	80.6	79.4	77.2	75.1	72.8	70.5	68.1	65.6	63.1	57.6	44.6	25.7
3	*****	66.1	65.8	64.8	63.1	61.3	59.5	57.6	55.6	53.6	51.5	47.0	36.4	21.0
4	*****	57.3	57.0	56.1	54.6	53.1	51.5	49.9	48.2	46.4	44.6	40.7	31.5	18.2
5	*****	51.2	51.0	50.2	48.9	47.5	46.1	44.6	43.1	41.5	39.9	36.4	28.2	16.3
6	*****	46.8	46.5	45.8	44.6	43.3	42.0	40.7	39.3	37.9	36.4	33.2	25.7	14.9
7	*****	43.3	43.1	42.4	41.3	40.1	38.9	37.7	36.4	35.1	33.7	30.8	23.8	13.8
8	*****	40.5	40.3	39.7	38.6	37.5	36.4	35.3	34.1	32.8	31.5	28.8	22.3	12.9
9	*****	38.2	38.0	37.4	36.4	35.4	34.3	33.2	32.1	30.9	29.7	27.1	21.0	12.1
10	*****	36.2	36.0	35.5	34.5	33.6	32.6	31.5	30.5	29.4	28.2	25.7	19.9	11.5
11	*****	34.5	34.4	33.8	32.9	32.0	31.1	30.1	29.0	28.0	26.9	24.5	19.0	11.0
12	*****	33.1	32.9	32.4	31.5	30.6	29.7	28.8	27.8	26.8	25.7	23.5	18.2	10.5
13	*****	*****	31.6	31.1	30.3	29.4	28.6	27.7	26.7	25.7	24.7	22.6	17.5	10.1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
125	*****	*****	*****	*****	*****	9.5	9.2	8.9	8.6	8.3	8.0	7.3	5.6	3.3
150	*****	*****	*****	*****	*****	8.7	8.4	8.1	7.9	7.6	7.3	6.6	5.1	3.0
200	*****	*****	*****	*****	*****	7.3	7.1	6.8	6.6	6.3	5.8	4.5	2.6	
250	*****	*****	*****	*****	*****	*****	*****	6.3	6.1	5.9	**5.6**	5.1	4.0	2.3
300	*****	*****	*****	*****	*****	*****	*****	**5.8**	5.6	5.4	5.1	4.7	3.6	2.1
350	*****	*****	*****	*****	*****	*****	*****	*****	5.1	5.0	4.8	4.4	3.4	1.9
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	4.6	4.5	4.1	3.2	1.8
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	4.2	3.8	3.0	1.7
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.6	2.8	1.6
750	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.3	1.3
1,000	******	****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.2

NOTE: For correct usage of these tables, please refer to the microdata documentation.

Example 4: Estimates of Ratios

Suppose that the user estimates that 297,735 persons in the SOW population in Quebec are employed in the public sector, while 427,088 persons in the SOW population in Ontario are employed in the public sector. The user is interested in comparing the estimate of Quebec versus that of Ontario in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of individuals in Quebec employed in the public sector. The denominator of the estimate ($\hat{X}_2$) is the number of individuals in Ontario employed in the public sector.

2) Refer to the coefficient of variation table for QUEBEC and ONTARIO.

3) The numerator of this ratio estimate is 297,735 - the estimate for Quebec. The figure closest to it is 300,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 5.8%.

4) The denominator of this ratio estimate is 427,088 - the estimate for Ontario. The figure closest to it is 450,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 5.0%

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

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

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

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0.058)^2 + (0.050)^2} \\ &= \sqrt{0.003364 + 0.0025} \\ &= 0.077 \end{aligned}$$

6) The obtained ratio of Quebec versus Ontario individuals employed in the public sector is 297,735 / 427,088 which is 0.697 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 7.7%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of individuals without Registered Retirement Savings Plans (RRSP), to individuals with RRSPs is 0.423 for Quebec, while it is 0.319 for Ontario. 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 Quebec ratio ($\hat{R}_1$) and the Ontario ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Quebec ratio is 5.6% and 5.1% for Ontario.

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 $\hat{d}$ = 0.423 – 0.319 = 0.104 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.423)(0.056)]^2 + [(0.319)(0.051)]^2} \\ &= \sqrt{(0.0005611) + (0.0002646)} \\ &= 0.029\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.029 / 0.104 = 0.279.

4) So the approximate coefficient of variation of the difference between the estimates is 27.9%. The difference between the estimates is considered marginal and Statistics Canada recommends this estimate be flagged with the letter E (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

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 persons aged 55 to 59 in Ontario who were working and never retired (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 82.3% (or expressed as a proportion 0.823)

t = 2

$\alpha_{\hat{x}}$ = 1.7% (0.017 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.823 – (2) (0.823) (0.017), 0.823 + (2) (0.823) (0.017)}

$CI_{\hat{x}}$ = {0.823 – 0.028, 0.823 + 0.028}

$CI_{\hat{x}}$ = {0.795, 0.851}

With 95% confidence it can be said that between 79.5% and 85.1% of persons aged 55 to 59 in Ontario were working and never retired.

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 two person households in Quebec with household income of $50,000 to $100,000 and the proportion of two person households in Ontario with household income of $50,000 to $100,000. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.032. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.398 - 0.405}{0.032} = \frac{-0.007}{0.032} = -0.219$$

Since t = -0.219 is between -2 and 2, then no conclusion about the difference can be made at the 5% 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 SOW 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 months without work following a job displacement would be greater than the coefficient of variation of the corresponding proportion of workers that had experienced a job displacement. 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 SOW2008_CVTabsE.pdf for the coefficient of variation tables.

11.0 Weighting

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

11.1 Weighting Procedures for the Labour Force Survey

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

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

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

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

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

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

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

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

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

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

11.2 Weighting Procedures for the Survey of Older Workers

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

1. Adjustment for computer-assisted personal interview (CAPI) and LFS non-response.

All LFS sample units appearing on the LFS Master file (TABS file) were assigned a survey weight. However, not all units on the TABS file were respondents in the month the data was collected. In some cases when a response could not be obtained for whatever reason, data for the month were imputed, often by carrying forward data from a previous month. Since these were actually cases of LFS non-response and since the SOW was to be selected from the LFS respondents, a separate non-response adjustment was made to account for these cases prior to the SOW weighting. The LFS non-response adjustment groups were used for this re-weighting step.

As well, LFS respondents for which data was collected by personal interview were excluded from the SOW sample. These sample units were grouped with LFS non-respondents and included in the above mentioned non-response adjustment.

2. Adjustment for rotation groups

The LFS sample is comprised of six rotation groups while the SOW sample uses only five of them. Therefore an adjustment of 1.2 is applied to the SOW weight to adjust for the missing rotation group.

3. Adjustment for sub-sampling

In selecting the sample for the SOW, all LFS households in the five SOW rotation groups with at least one person in the target population for the SOW were contacted for the survey. If there was more than one person in the household belonging to the target population, then one person was selected at random for the SOW sample. To account for sub-sampling, the SOW weight is multiplied by the number of household members belonging to the target population.

4. Adjustment for non-response

To adjust for non-response to the SOW, logistic regression was used to estimate the expected probability of response for each sample unit. Modeling was done within province. To form response groups for the adjustment, the sample file was sorted by probability of response and then divided into deciles giving ten response adjustment groups for each province. Within the adjustment groups, the weight adjustment factor was computed as the ratio of the sum of the weights for all sample units to the sum of the weights for all respondents.

5. Adjustment for calibration to projected population totals

A calibration of non-response adjusted weights is then performed. Non-response adjusted SOW weights for SOW respondents are combined with LFS sub-weights for LFS respondents aged 50 to 75 and not in the SOW target population. A calibration is then done adjusting weights so that the weights sum to known demographic estimates for the 50 to 75 year old age group. Calibration had to be done including the LFS non-SOW sample since the demographic estimates were only available for the whole population of 50 to 75 year olds and not for the SOW target population.

The resulting weight WTPM is the final weight which appears on the SOW master microdata file and WTPP for the SOW public use microdata file (PUMF).

12.0 Questionnaires

12.1 The Labour Force Survey Questionnaire

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

12.2 The Survey of Older Workers Questionnaire

The Survey of Older Workers (SOW) questionnaire was used in October and November of 2008 to collect the information for the supplementary survey. The file SOW2008_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See SOW2008_CdBk.pdf for the record layout with univariate frequencies.