Microdata User Guide

Households and the Environment Survey

2006

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Households and the Environment Survey (HES) was conducted by Statistics Canada in February, March and April 2006. 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

Environment Accounts and Statistics Division
Telephone: 613-951-0297
Fax: 613-951-0634
E-mail: environ@statcan.ca

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

2.0 Background

The Households and the Environment Survey (HES) was conducted in February, March and April of 2006 as a supplement to the Labour Force Survey. The survey was designed to specifically address the needs of its funding source the Canadian Environmental Sustainability Indicators (CESI) project, a joint venture between Statistics Canada, Environment Canada and Health Canada. The CESI project reports annually on air quality, water quality and greenhouse gas (GHG) emissions in Canada using indicators to identify areas of importance to Canadians and monitor progress.

The HES was first conducted in 1991 and again in 1994. The 2006 survey offers an expanded view on household behaviours that relate to the environment but allows for comparisons with the 1994 survey for some indicators.

The target population for the HES consisted of households in Canada, excluding households in which no member is 18 years of age or older. Also excluded were households located in the Yukon, Northwest Territories and Nunavut, households located on Indian reserves and on military bases, and households consisting entirely of full-time members of the Canadian Armed Forces. For a subset of questions, the survey targeted adults 18 years of age or older living in households that were included in the survey's main target population. The survey, therefore, provides two different units of analysis: the household for most questions, and the person for a limited number of questions relating to modes of transportation used to travel to work.

3.0 Objectives

The objective of the Households and the Environment Survey (HES) is to measure the behaviours and practices of households that relate to the environment in terms of their impact on the quality of the air, water and soils as well as contributions to greenhouse gas emissions. Specifically, the following topics were addressed in the 2006 HES:

- Water quality concerns of households
- Consumption and conservation of water
- Energy use and home heating and cooling
- Use of gasoline powered equipment
- Pesticide and fertilizer use on lawns and gardens
- Recycling, composting and waste disposal practices
- Impacts of air and water quality on households
- Motor vehicle use and transportation decisions for travel to work

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 Households and the Environment Survey (HES) 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.

Dwelling
A dwelling is defined as any set of living quarters that is structurally separate and has a private entrance outside the building or from a common hall or stairway inside the building.

Types of dwellings
- **Single detached** – A structure with one dwelling only, separated by open space from all other structures except its own garage or shed.
- **Double** – A dwelling joined to only one other dwelling, separated from it by a wall extending from ground to roof.
- **Row or terrace** – A dwelling unit in a row of three or more dwellings, sharing common walls extending from ground to roof in which there are no other dwellings either above or below.
- **Duplex** – Two dwellings situated one above the other, not attached to any other structure and surrounded on all sides by open space.
- **Low-rise apartment** – Dwellings within triplexes, quadruplexes, and apartment buildings of fewer than five stories.
- **High-rise apartment** – Separate dwellings within a residential structure of five or more stories.
- **Mobile homes** – A movable dwelling designed and constructed to be transported (by road) on its own chassis to a site, and placed on a temporary foundation such as block posts or a prepared pad.

Household
A household is defined as any person or a group of persons living in a dwelling. A household may consist of any combination of: one person living alone, one or more families, a group of people who are not related but who share the same dwelling.

4.2 Households and the Environment Survey Concepts and Definitions

Consumption and conservation of water
"Canadians are concerned about how the environment affects their health, thus about the quality of the water they drink"[3]. Since public perception, as a determining factor driving public policy, can be as persuasive as empirically-based evidence, it is important to understand how Canadians perceive the quality of their drinking water supply and the behaviours they are exhibiting that may reflect their concerns.

These behaviours are measured in the HES through the purchases of bottled water or the use of water filters, and the reasons for making these purchases – for example, concerns about bacterial contamination. Analysis can be carried out to gain insight into the household characteristics of those that do and do not exhibit behaviours that may indicate uncertainties about the quality of their drinkable water.

Aside from drinking water issues, another important theme involves the water conservation practices of households. Water scarcity is an emerging issue for many Canadians and this concern may be exacerbated by climatic changes. Many regions of Canada have experienced drought or near-drought conditions which has led to regulatory responses by municipal authorities (e.g., water use restrictions) and/or the voluntary adoption of water conservation measures by households.

The HES provides reliable information on household practices such as lawn watering and the use of low-flow showerheads, which can be used to determine levels of public participation in water conservation.

Metals and minerals can be any of the following: iron, sulphur, cadmium, zinc, manganese, lead, mercury, arsenic.

Bacteria can be any of the following: E. coli, coliform, fecal matter, giardia, crypto, cryptosporidium, parasites, protozoa, shigellosis.

Chemicals or other pollutants can be any of the following: chlorine, bromine, pesticides, oil, gasoline, diesel fuel, heating oil, fluoride, nitrate, trichloroethylene (TCE), polycyclic aromatic hydrocarbons (PAH), fire retardants.

Holding tanks are septic tanks that do not have a weeping tile system and which must be pumped out on a regular basis.

A **communal septic system** is a private or public septic system that serves more than one household but is not a part of a municipal sewer system. These are common in places like trailer parks or neighbourhoods where there is not a high enough housing density to warrant full sewage services.

Low flow showerheads are able to regulate the flow of water.

Low volume toilets use a lower volume of water than regular toilets. Usually these toilets use 6 litres as opposed to 12 litres of water per flush.

A rain barrel is a container used to collect and store rainwater. It is usually placed below the downspout of a roof gutter. The collected water is usually used to water the landscape.

[3] Sharing Environmental Decisions: Executive Summary and Recommendations. Final report of the Task Force on a Canadian Information System for the Environment. October 2001, Ottawa.

A cistern is an artificial reservoir for storing liquids; especially an underground tank for storing rainwater.

Energy use and home heating

The choices regarding what kind of energy a household uses to heat a home, whether the temperature is regulated and if any energy efficient electrical devices are used within the home are all decisions that affect the household's contributions to greenhouse gas (GHG) emissions. When households demand more electricity, for example, this demand translates to the need for a greater energy output by coal-fired electrical utilities – which are the largest single contributors to GHGs.

Indirectly, the decisions also indicate to what degree the household has bought into the concept and need for energy conservation. The HES measures not only the types of energy used (and in the case of wood, the quantity), but those behaviours that indicate whether Canadian households are behaving in a sustainable way regarding energy use.

A **forced air natural gas furnace** is a heating system using a set of ducts and vents to circulate air heated by combusting natural gas.

A **forced air oil furnace** is a heating system using a set of ducts and vents to circulate air that has been heated by combusting oil.

A **forced air electric furnace** is a heating system using a set of ducts and vents to circulate air that has been heated by electrical current.

A **forced air hot water system** is a heating system using a set of ducts and vents to circulate air that has been heated by hot water.

Hot water radiators are metal structures or pieces of equipment used to heat a room by emitting heat from hot water or steam that circulates through it.

Electric baseboards are heating systems attached to the wall near the floor where elements heat up through use of electrical current. A set of electric baseboards are controlled with independent thermostats - usually one per room.

Other electric heating is heat produced through electrical current that is delivered through an appliance or other means excluding forced air or baseboards.

Wood stove or wood fireplace is where you can have a domestic fire using wood as fuel. A chimney is used for ventilation of smoke and excess heat.

A **face cord** of wood also known as a stove cord measures 8 feet long by 4 feet high and is usually 12 or 16 inches wide.

A **full cord** of wood also known as a bush cord measures 8 feet long by 4 feet high and is usually 4 feet wide. There are either 4 (using 12 inch pieces) or 3 (using 16 inch pieces) face cords in a full or bush cord.

A **central air conditioning system** is part of the home's central heating system and distributes cool air through the home's ductwork as opposed to stand alone air conditioning units that are usually seen in windows and are used to cool a specific part of the home.

Gasoline powered equipment use

Internal combustion engines that power small devices such as lawnmowers or those that power equipment such as boats and snowmobiles all emit GHGs - which are believed to be a main

contributor to climate change, as well as fine particulates – which adversely affect air quality. By measuring the extent of the use of these devices, we are establishing a baseline that can be compared to future data to see if the use of these devices are increasing or decreasing.

A **grass trimmer (weed eater)** is a gasoline powered device that trims weeds though the use of a rapidly spinning plastic cord or circular saw.

A **leaf blower** is a gasoline powered device that emits a strong air current that is intended to blow leaves off of one's lawn.

Pesticide and fertilizer use

Since the Supreme Court of Canada decision that granted authority to the municipality of Hudson, Quebec to ban the residential use of pesticides, there has been increasing interest at all levels of government regarding the extent and nature of pesticide use by Canadian households. This increased interest was also reflected in the May 2000 report on *Pesticides* by the Standing Committee on Environment and Sustainable Development.

The report published United States data on the use of pesticides in urban communities, demonstrating the need for Canadian data on urban pesticide use. Health Canada's Pest Management Regulatory Agency *Action Plan on Urban Use of Pesticides* has focused on the indoor and outdoor urban use of pesticides, adoption by households of healthy lawn care practices and landscape services.

The usage of fertilizers and pesticides, or fungicides by households is measured in the 2006 HES. By analysing these data with selected household characteristics, policymakers can then use this information to better inform targeted public awareness and information campaigns.

Chemical fertilizers are chemicals given to plants with the intention of promoting growth. They are usually applied either via the soil or by spraying.

Weed killers, pesticides, fungicides are a substance or mixture of substances intended for preventing, destroying, repelling, or mitigating any pest. It is also any substance or mixture of substances intended to regulate plant or leaf growth. A pest can be an insect, an unwanted plant such as a weed or a fungal growth on a plant or tree.

Recycling, composting and waste disposal practices

Disposal and recycling of post-consumer waste continues to be an important environmental issue, especially in urban centers. Canadian households generate approximately one third of solid waste by weight. As we move into a "knowledge-based" economy, the amount of "high-tech" waste has increased and represents new waste management problems that were not covered in the previous HES. In addition, there is a need for better data on household practices regarding composting, participation in product stewardship programs, knowledge of and participation in hazardous waste disposal programs. For instance, the Extended Producer Responsibility (EPR) program is a policy approach for which producers accept significant responsibility (financial and/or physical) for the treatment or disposal of post-consumer products like beer bottles and cans, beverage containers, tires, used paint, used oil, etc. Assigning such responsibility could provide incentives to prevent wastes at source, promote product design for the environment, and support the achievement of public recycling and materials management goals.

Participation in recycling initiatives

Governments are finding it difficult to gauge the level of participation in various waste diversion initiatives concerning certain materials such as used paint, expired medication and used batteries. Scores of such initiatives are underway across Canada and there is a perception that the level of success for these programs is high. However, data to accurately measure the *extent* of household involvement in these programs do not exist. Thus, the

participation rates in these programs are a data gap to be filled by the HES. As in 1991 and 1994, the 2006 HES continues to ask questions on access and use of recycling programs.

Household computer and technological waste

The need for better data on household practices regarding how households deal with their technological waste is an issue that arose through contacts with provincial officials and has been on the agenda at the Federal level[4]. Concerns about the levels of toxic substances (e.g.., lead, mercury and dioxins) that can be found in these wastes are heightening as policymakers examine methods to mitigate the impact of these materials on the environment. The HES measures whether households are disposing of technological waste in proper disposal facilities or whether they are including it in the "regular" garbage or are perhaps simply unaware of what to do with these materials and are storing them in basements and garages.

Composting practices

Usage by households of backyard composters and/or curb side organic pick-up represents another data gap in waste statistics. There is a high level of interest from all levels of government and NGOs (non-government organizations) regarding household usage of backyard composters and the socio-economic characteristics of users and non-users of backyard composters and centralized composting programs

A **recycling program** can be either:

- A collection system for recyclable materials such as paper, plastics, metals, and glass. This system can be either municipally or privately operated.
- A system whereby residents and/or businesses take their recyclable materials to a central depot or drop-off centre.

Composting involves the separation of kitchen waste (includes food scraps, coffee grinds, eggshells, etc.) and or yard waste (includes leaves, plants or grass clippings) from the rest of your household garbage. The separated materials can be:

- Put in a compost bin, compost pile or your garden
- Picked up by your city, town, municipality or a private company; OR
- Taken to a depot or drop off centre

Air and water quality

The quality of the air and the bodies of water (lakes, rivers) we use for recreational purposes may perhaps influence how Canadians behave in their day to day life activities. A failure to alter our behaviour can impact the quality of life and health for all Canadians. The HES measures, for example, how smog advisories or swimming restrictions have influenced people's activities.

Smog is actually a combination of the words smoke and fog. Smog is the most visible form of air pollution. It is a brownish-yellow hazy cloud caused when heat and sunlight react with various pollutants emitted from industry, cars, pesticides and oil based home products. Smog is a year-round problem but most smog watches and alerts occur from May to September, especially on hot days.

An **air quality advisory** is a government warning issued when air emissions - substances discharged into the air by motor vehicle engines and other sources - contain pollutants that are at potentially harmful levels. Alerts are based on the air quality index.

Transportation decisions

[4] Information Technology (IT) and Telecommunication Waste in Canada. EnvirosRIS. Prepared for Environment Canada, National Office of Pollution Prevention. October 2000, Ottawa.

Emissions from automobiles are a significant component of greenhouse gas emissions in Canada, and thus of climate change. Policies that seek to educate the public about the impact of their commuting choices as well as policies around public transportation issues require reliable information about the behaviours and motivations of Canadians regarding their commuting choices. The HES measures the commuting choices as well as length of time spent commuting and the distances traveled by individuals.

Ethanol blended fuel is gasoline blended with ethanol. Ethanol is the alcohol produced from the starch portion of corn. This gasoline is usually marketed as more "environmentally friendly" than other types of gasoline.

5.0 Survey Methodology

The Households and the Environment Survey (HES) was administered in February, March and April 2006 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.[5] Sections 5.5 and 5.6 describe how the HES departed from the basic LFS design in February 2006.

5.1 Population Coverage

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

5.2 Sample Design

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

5.2.1 Primary Stratification

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

[5] 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 Households and the Environment Survey

The HES sample was selected from the February 2006 LFS sample. Dwellings in the LFS rotations were selected in each CMA (28 domains) and the non-CMA portion of each province (10 domains) to yield estimates of proportions of approximately 10% with a coefficient of variation of 16.5%. The entire sample of LFS responding households in a rotation were selected within a particular domain in order to respect the design of the LFS. Households that responded to the LFS using the computer-assisted personal interview (CAPI) application were excluded from the sample and coded as non-response. Supplementary rotations in CMA and non-CMA domains were also selected where warranted to improve the quality of estimates in various drainage basins and sub-basins.

The collection of the HES sample was distributed over three months starting in February in order to minimize the impact on the LFS and reduce respondent burden. The HES was conducted as a supplement to the LFS in February, March and April as follows:

- parts of rotation groups 3 and 4 were interviewed in the February LFS
- parts of rotation groups 1 and 5 were interviewed in the March LFS
- parts of rotation groups 2 and 6 were interviewed in the April LFS.

For the HES, the coverage of the LFS was modified to include members of the household aged 18 and over.

5.6 Sample Size by Province for the Households and the Environment Survey

The following table shows the number of households in the LFS sampled rotations that were included in the sample for the HES supplement. This table excludes households which were non-respondents to the LFS and households that responded to the LFS in person using a CAPI method.

Sample size by month of collection

Province	February 2006	March 2006	April 2006	Total
Newfoundland and Labrador	295	580	588	1,463
Prince Edward Island	-	410	389	799
Nova Scotia	243	852	884	1,979
New Brunswick	115	447	458	1,020
Quebec	1,512	2,548	2,547	6,607
Ontario	2,472	4,557	4,500	11,529
Manitoba	-	1,088	563	1,651
Saskatchewan	274	1,134	890	2,298
Alberta	858	1,725	1,672	4,255
British Columbia	1,035	1,903	1,892	4,830
Canada	**6,804**	**15,244**	**14,383**	**36,431**

6.0 Data Collection

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

6.1 Interviewing for the Labour Force Survey

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

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

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

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

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

6.2 Supervision and Quality Control

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

6.3 Non-response to the Labour Force Survey

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

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

6.4 Data Collection Modifications for the Households and the Environment Survey

Upon completion of the Labour Force Survey interview, the interviewer asked to speak to the person most knowledgeable about the household activities that relate to the environment. If the respondent did not agree to continue at that time, the interviewer arranged for a convenient time to phone back. Collection for the HES was completed within one week of the end of the LFS collection period in order to minimize the number of contacts with LFS responding households.

During the interview, some questions were addressed to a randomly selected person 18 years of age or older. These questions were answered by proxy when the selected subject was not the HES respondent.

6.5 Non-response to the Households and the Environment Survey

As a voluntary supplement survey to the LFS, little effort was made in attempting to convert refusal cases. Non-response was nevertheless very low. Interviewers noted that generally respondents were happy to answer and interested in the survey. The majority of the refusals for the HES were also refusals and refusal conversions to the LFS.

7.0 Data Processing

The main output of the Households and the Environment Survey (HES) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

Since the HES collected information relating to both the household and a randomly selected person 18 years of age or older, the results have been stored in two separate data files: the HOUSEHOLD file and the PERSON file.

The **HOUSEHOLD** microdata file contains data from the following sections as well as LFS variables related to the household:

- WA Water
- EH Energy Use and Home Heating
- GP Gasoline Powered Equipment
- FP Fertilizer and Pesticide Use
- RC Recycling
- CP Composting
- AQ Air and Water Quality
- TD Transportation Decisions - questions that relate to households
- HD Household Demographics - questions that relate to household income

The **PERSON** microdata file contains data from the following sections as well as LFS variables related to the randomly selected person and the household:

- WA Water
- EH Energy Use and Home Heating
- GP Gasoline Powered Equipment
- FP Fertilizer and Pesticide Use
- RC Recycling
- CP Composting
- AQ Air and Water Quality
- TD Transportation Decisions - questions which are specific to respondents who worked outside the home
- HD Household Demographics - questions that relate to household income

7.1 Data Capture

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

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

7.2 Editing

The first stage of survey processing undertaken at head office was the replacement of any "out-of-range" values on the data file with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions that did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions that should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

This was followed by a series of edits to ensure consistency in the responses for a household.

7.3 Coding of Open-ended and Numeric Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. Two open-ended questions were included in the survey and there were six questions which allowed for a range of numeric values to be entered. These questions required coding for inclusion on the HES data file.

The open-ended questions related to other types of treatments for drinking water and for other measurements of wood to be entered.

The second type of coding performed was for questions which allow for numeric values to be entered. These numeric values were first reviewed for outliers and then grouped into ranges. Specifically, the numeric questions in the HES were: number of cords of wood, temperature of the home when awake and asleep, volume of waste which is composted, number of vehicles owned or leased for personal use and the total household income from all sources.

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. The 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 a minimum set of answers but does not finish the interview. These records can be handled either as complete non-response or multiple item non-response.

In the case of the HES, item non-response was treated as not-stated. Partial non-response was treated as multiple item non-response since less than 1% of respondents only partially completed the interview.

7.5 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. This was done by using one variable or a combination of variables. The following is a list of the derived variables for the HES. The first table is a list of the derived variables that apply strictly to the HES followed by derived variables based on the LFS variables and pertaining to the HOUSEHOLD file and the PERSON file.

HES Derived Variables

WAD03	Indication if any treatment is applied to the drinking water in the household
WAD09	Indication if any maintenance is done to the septic system
WADREDUC	Indication if any of the noted devices are used by the household to conserve or reduce consumption of water
EHD04	Face cords burned in a heating season
EHD10	During the heating season, at what temperature do you normally keep your home when you are there and awake?(Celsius)
EHD11	During the heating season, at what temperature do you normally keep your home when you are asleep?(Celsius)
GPDEQUIP	Usage of snow blower, lawn mower, weed eater, or leaf blower
RCDACCES	Access to any recycling program such as glass, paper, plastics, or metal cans
RCDPROGR	Usage of any recycling program
RCDDISPO	Usage of any waste disposal practice for paint, medications, batteries or technology equipment
AQDRESTR	Smog advisories or swimming restrictions influenced people's behavior and activities
HDD02	Household income
TDD08**	What is your main mode of transportation to work during the colder months?
TDD15**	What is your main mode of transportation to work during the warmer months?

** Variables found on the person file only

LFS Derived Variables

HHSIZE	Number of people in the household
NUMFAM	Number of economic families in the household
HHEDUCLV	Highest level of education ever completed by any member of the household
HHEMPL	Total number of employed persons in all the economic families in the household
HHNWKE	Total number of employees in all the economic families in the household
HHUNEM	Total number of unemployed persons in all the economic families in the household
HHWKE	Sum of usual weekly earnings for all employees in all the economic families in the household
DRAINAGE	Sub-basin description
DRAINSB1	Sub-basin to oceans or bays
DRAINSB2	Sub-basin to oceans or bays – high level
CMA	Census metropolitan areas (CMA) and non-CMAs of each province
FAMIDMN	ID for the main economic family in the household (i.e. the family with the most members, or if all the families are of equal size then first FAMID (usually “A”)

EFAMSZMN	Number of individuals in the main economic family
EFAMTYMN	Type of economic family for the main economic family
HHAG0005	Number of persons aged 0 to 5 in the household
HHAG0612	Number of persons aged 6 to 12 in the household
HHAG1315	Number of persons aged 13 to 15 in the household
HHAG1617	Number of persons aged 16 to 17 in the household
HHAG1819	Number of persons aged 18 to 19 in the household
HHAG2024	Number of persons aged 20 to 24 in the household
HHAG2534	Number of persons aged 25 to 34 in the household
HHAG3544	Number of persons aged 35 to 44 in the household
HHAG4554	Number of persons aged 45 to 54 in the household
HHAG5564	Number of persons aged 55 to 64 in the household
HHAG65PL	Number of persons aged 65 and over in the household
PDAGE **	Age of respondent
PMARSTAT **	Grouped marital status of respondent

** Variables found on the person file only

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 households who treat their drinking water is to be estimated, it is done by selecting the records referring to those households 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 output of the 2006 HES has been stored in two separate data files: the HOUSEHOLD file and the PERSON file, a PUMF is available only for the HOUSEHOLD file.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Households and the Environment Survey (HES).

Province	LFS Selected Households	LFS Response Rate *	HES Selected Households **	HES Sample for Collection	HES Responding Households	HES Collection Response Rate ***	HES Response Rate ****	HES Overall Response Rate *****
		%				%		
Newfoundland and Labrador	1,605	92.5	1,567	1,463	1,197	81.8	76.4	74.6
Prince Edward Island	943	92.7	914	799	649	81.2	71.0	68.8
Nova Scotia	2,187	92.5	2,105	1,979	1,649	83.3	78.3	75.4
New Brunswick	1,176	90.6	1,134	1,020	830	81.4	73.2	70.6
Quebec	7,309	91.5	7,074	6,607	5,416	82.0	76.6	74.1
Ontario	12,955	91.1	12,543	11,529	8,827	76.6	70.4	68.1
Manitoba	1,821	91.0	1,757	1,651	1,288	78.0	73.3	70.7
Saskatchewan	2,481	91.0	2,403	2,298	1,791	77.9	74.5	72.2
Alberta	4,506	90.8	4,383	4,255	3,155	74.1	72.0	70.0
British Columbia	5,363	91.0	5,194	4,830	3,532	73.1	68.0	65.9
Canada	**40,346**	**91.3**	**39,074**	**36,431**	**28,334**	**77.8**	**72.5**	**70.2**

Note: The LFS counts are in terms of households while the HES counts are in terms of households as well as selected individuals within households (only one individual was selected per household). The HES overall response rate is based on all LFS household records. The LFS responding households include respondents carried forward from the previous month.

* The LFS response rate is the number of LFS responding households as a percentage of the number of LFS selected households in the rotation groups (panels) in scope for each of the three months of collection for the HES.

** The HES selected households that were interviewed using a computer-assisted personal interview (CAPI) application during the LFS were not included in the HES sample for collection and were treated as non-response for the HES.

*** The HES collection response rate is the number of HES responding households as a percentage of the HES sample for collection.

**** The HES response rate is the number of HES responding households as a percentage of the number of HES selected households in the month of collection. The HES selected households include all computer-assisted telephone interview (CATI) and CAPI households in the rotation groups (panels) in scope for each of the three months of collection for the HES.

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

8.2 Survey Errors

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

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

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

8.2.1 The Frame

Because the HES was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the HES frame. The quality of the sampling variables in the frame was very high. The HES sample, which was selected from the February 2006 LFS sample file, consisted of at least two rotation groups within each census metropolitan area during the three months of collection. The criteria used for the HES selection (like rotation group) were not missing for any LFS records.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the HES frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data.

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

8.2.2 Data Collection

Interviewer training consisted of reading the HES Procedures Manual and Interviewer's Manual, practicing with the HES training cases on the computer and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers collected the HES information after the

LFS information was collected. The interviews were conducted during the LFS survey week and one week following in February, March and April, 2006.

Rules for generating the HES sample
The HES sample was selected for all three months of collection from the February LFS sample of responding households. The sample for the February, March and April collection periods were verified before production each month to include only responding households who completed the LFS by telephone. Once verification was completed, a list of the sampled units for the given month was added to the monthly sample file. The application was programmed to read the attached list and generate a HES component if the identifier matched the one on the list.

8.2.3 Data Processing

During processing of the data, six HES responding household records did not match to corresponding records in the LFS. Thus they were coded as out-of-scope and were dropped from further processing. When supplementary survey records do not match to host survey records they must be dropped since a weight cannot be derived for them.

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

8.2.4 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of individuals who/households that responded to the survey to compensate for those who/that did not respond.

For the LFS 2,643 cases completed the survey in person using a computer-assisted personal interview (CAPI) application and were treated as non-response for the HES.

In most cases, item 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. Values were not imputed when these were missing and were coded to “not-stated”.

Partial non-response occurs when the interview is started but not completed for various reasons. In the case of the HES, less than 1% of interviews were started but not completed and the missed questions were treated as multiple item non-response and coded to “not-stated”.

8.2.5 Measurement of Sampling Error

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

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

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

For example, suppose that, based upon the survey results, one estimates that 34.9% of households had a lawn and used chemical fertilizers in 2005 and this estimate is found to have a standard error of 0.0051. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0051}{0.349}\right) X\ 100\,\% = 1.46\,\%$$

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.

There are two microdata files for the Households and the Environment Survey **(**HES), the HOUSEHOLD file and the PERSON file. The survey collected information relating to both the household and to a randomly selected person 18 years of age or older for questions relating to the mode of travel to work.

The **HOUSEHOLD** microdata file should be used for analysis and estimation of the majority of the content of the survey where the unit of measurement was the household.

The **PERSON** microdata file should be used for analysis of the questions in the Transportation Decisions section that relate to travel to work. Analysis and estimation of other variables in the person file should be avoided and used only as a characteristic of the variables where the unit of measurement was the randomly selected person. For example, an analyst could evaluate if adults who use public transportation to travel to work are more or less likely to live in households that exhibit other environmentally friendly behaviours such as conserving water and energy, recycling and composting.

The examples provided in Chapter 9.0 and 10.0 are based on the household file.

9.1 Rounding Guidelines

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

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

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

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

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

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released

which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the HES was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper survey weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

9.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the HES 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 HES.

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 households that have access to a glass recycling program or the proportion of households that primarily drink bottled water are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Does your household have access to a recycling program for bottles?
R: Yes / No

Q: What type of water does your household primarily drink at home?
R: Tap water / Bottled water / Both / Other

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 motor vehicles owned or leased by households. The numerator is an estimate of the total number of motor

vehicles owned or leased by households, and its denominator is the number of households with a vehicle.

Examples of Quantitative Questions:

Q: How many motor vehicles are owned or leased for personal use by your household?
R: |_|_| motor vehicles

Q: How many months a year do you compost you kitchen waste?
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 amount of wood burned in a heating season for households that use wood as their main energy source for heating, multiply the value reported in derived variable EHD04 (amount of wood in face cords) by the final weight for the record, then sum this value over all records with EH_Q01 = 8 (main type of heating equipment is a wood stove or wood fireplace.)

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 amount of wood burned in a heating season for households that use wood as their main energy source for heating,

a) estimate the total amount of wood burned ($\hat{X}$) as described above,
b) estimate the number of households were the main type of heating equipment is a wood stove or wood fireplace ($\hat{Y}$) by summing the final weights of all records with EH_Q01 = 8, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.4 Guidelines for Statistical Analysis

The HES 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 Quebec households is required. The steps to rescale the weights are as follows:

1) select all households from the file in PROV = 24, Quebec;

2) calculate the AVERAGE weight for these records by summing the original household weights from the microdata file for these records and then dividing by the number of households in PROV = 24;

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

4) perform the analysis for these households using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the HES, 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 Households and the Environment Survey

The following tables 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 households possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding

quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explains the correct procedure to be used for ratios.

Table of Release Cut-offs – Household File

Province and Region	Acceptable CV 0.0% to 16.5%	Marginal CV 16.6% to 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	8,000 & over	2,000 to < 8,000	under 2,000
Prince Edward Island	3,600 & over	900 to < 3,600	under 900
Nova Scotia	11,000 & over	2,800 to < 11,000	under 2,800
New Brunswick	21,100 & over	5,500 to < 21,100	under 5,500
Quebec	87,500 & over	21,900 to < 87,500	under 21,900
Ontario	40,600 & over	10,000 to < 40,600	under 10,000
Manitoba	19,000 & over	4,800 to < 19,000	under 4,800
Saskatchewan	12,600 & over	3,200 to < 12,600	under 3,200
Alberta	26,800 & over	6,700 to < 26,800	under 6,700
British Columbia	29,600 & over	7,400 to < 29,600	under 7,400
Atlantic Provinces	12,900 & over	3,200 to < 12,900	under 3,200
Prairie Provinces	21,900 & over	5,400 to < 21,900	under 5,400
Canada	**42,400 & over**	**10,400 to < 42,400**	**under 10,400**

Table of Release Cut-offs – Person File

Province and Region	Acceptable CV 0.0% to 16.5%	Marginal CV 16.6% to 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	18,000 & over	4,600 to < 18,000	under 4,600
Prince Edward Island	7,700 & over	2,000 to < 7,700	under 2,000
Nova Scotia	23,700 & over	6,000 to < 23,700	under 6,000
New Brunswick	41,200 & over	10,700 to < 41,200	under 10,700
Quebec	169,300 & over	42,500 to < 169,300	under 42,500
Ontario	92,800 & over	22,900 to < 92,800	under 22,900
Manitoba	37,600 & over	9,600 to < 37,600	under 9,600
Saskatchewan	25,000 & over	6,300 to < 25,000	under 6,300
Alberta	59,900 & over	15,000 to < 59,900	under 15,000
British Columbia	68,400 & over	17,000 to < 68,400	under 17,000
Atlantic Provinces	27,700 & over	6,900 to < 27,700	under 6,900
Prairie Provinces	47,600 & over	11,800 to < 47,600	under 11,800
Canada	**90,800 & over**	**22,300 to < 90,800**	**under 22,300**

10.0 Approximate Sampling Variability Tables

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

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

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province and region, which were used to produce the Approximate Sampling Variability Tables for the Households and the Environment Survey (HES) household file.

Household File

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.37	1,197	199,585
Prince Edward Island	1.29	649	53,581
Nova Scotia	1.35	1,649	377,202
New Brunswick	1.72	830	299,063
Quebec	4.12	5,416	3,223,408
Ontario	2.11	8,827	4,678,298
Manitoba	1.56	1,288	446,916
Saskatchewan	1.69	1,791	377,757
Alberta	1.91	3,155	1,234,693
British Columbia	1.73	3,532	1,678,037
Atlantic Provinces	1.67	4,325	929,431
Prairie Provinces	1.83	6,234	2,059,366
Canada	**2.62**	**28,334**	**12,568,540**

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province and region, which were used to produce the Approximate Sampling Variability Tables for the HES person file.

Person File

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.51	1,197	408,607
Prince Edward Island	1.39	649	105,839
Nova Scotia	1.52	1,649	725,255
New Brunswick	1.72	830	582,793
Quebec	4.33	5,416	5,940,869
Ontario	2.33	8,827	9,671,592
Manitoba	1.64	1,288	843,107
Saskatchewan	1.79	1,791	706,325
Alberta	2.14	3,155	2,465,540
British Columbia	2.02	3,532	3,326,176
Atlantic Provinces	1.82	4,325	1,822,494
Prairie Provinces	2.04	6,234	4,014,972
Canada	**2.84**	**28,334**	**24,776,103**

Because the design was adjusted in order to gain precision at the sub-basin level, Quebec was expected to have a higher design effect.

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

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

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

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

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

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

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

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of households that had a lawn (WA_Q16) and used chemical fertilizers in 2005 (FP_Q01) is more reliable than the estimated number of households that had a lawn and used chemical fertilizers in 2005. (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 province) 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 households that had a lawn and the numerator is the number of households that had a lawn and used chemical fertilizers in 2005.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of households that own their dwelling, had a lawn and used chemical fertilizers in 2005 as compared to the number of households that rent, had a lawn and used chemical fertilizers in 2005 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 HES 2006 are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

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

Suppose that a user estimates that 8,375,473 households had a lawn in 2005. 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 8,375,473 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 8,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 0.7%.

4) So the approximate coefficient of variation of the estimate is 0.7 %. The finding that 8,375,473 (to be rounded according to the rounding guidelines in Section 9.1) households had a lawn in 2005 is publishable with no qualifications.

Households and the Environment Survey, 2006 - Household File

Approximate Sampling Variability Tables - Canada

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE										
	0.1%	**1.0%**	**…**	**15.0%**	**20.0%**	**25.0%**	**30.0%**	**35.0%**	**40.0%**	**50.0%**	**70.0%**	**90.0%**
1	107.6	107.1		99.3	96.3	93.3	90.1	86.8	83.4	76.1	59.0	34.1
2	76.1	75.8		70.2	68.1	65.9	63.7	61.4	59.0	53.8	41.7	24.1
3	62.1	61.9		57.3	55.6	53.8	52.0	50.1	48.2	44.0	34.1	19.7
4	53.8	53.6		49.6	48.2	46.6	45.0	43.4	41.7	38.1	29.5	17.0
5	48.1	47.9		44.4	43.1	41.7	40.3	38.8	37.3	34.1	26.4	15.2
6	43.9	43.7		40.5	39.3	38.1	36.8	35.4	34.1	31.1	24.1	13.9
7	40.7	40.5		37.5	36.4	35.2	34.1	32.8	31.5	28.8	22.3	12.9
8	38.1	37.9		35.1	34.1	33.0	31.9	30.7	29.5	26.9	20.9	12.0
9	35.9	35.7		33.1	32.1	31.1	30.0	28.9	27.8	25.4	19.7	11.4
10	34.0	33.9		31.4	30.5	29.5	28.5	27.5	26.4	24.1	18.7	10.8
11	32.5	32.3		29.9	29.0	28.1	27.2	26.2	25.1	23.0	17.8	10.3
12	31.1	30.9		28.7	27.8	26.9	26.0	25.1	24.1	22.0	17.0	9.8
13	*****	29.7		27.5	26.7	25.9	25.0	24.1	23.1	21.1	16.4	9.4
14	*****	28.6		26.5	25.7	24.9	24.1	23.2	22.3	20.4	15.8	9.1
15	*****	27.7		25.6	24.9	24.1	23.3	22.4	21.5	19.7	15.2	8.8
16	*****	26.8		24.8	24.1	23.3	22.5	21.7	20.9	19.0	14.7	8.5
125	*****	9.6		8.9	8.6	8.3	8.1	7.8	7.5	6.8	5.3	3.0
150	*****	*****		8.1	7.9	7.6	7.4	7.1	6.8	6.2	4.8	2.8
200	*****	*****		7.0	6.8	6.6	6.4	6.1	5.9	5.4	4.2	2.4
250	*****	*****		6.3	6.1	5.9	5.7	5.5	5.3	4.8	3.7	2.2
300	*****	*****		5.7	5.6	5.4	5.2	5.0	4.8	4.4	3.4	2.0
350	*****	*****		5.3	5.1	5.0	4.8	4.6	4.5	4.1	3.2	1.8
400	*****	*****		5.0	4.8	4.7	4.5	4.3	4.2	3.8	2.9	1.7
450	*****	*****		4.7	4.5	4.4	4.2	4.1	3.9	3.6	2.8	1.6
500	*****	*****		4.4	4.3	4.2	4.0	3.9	3.7	3.4	2.6	1.5
750	*****	*****		3.6	3.5	3.4	3.3	3.2	3.0	2.8	2.2	1.2
1,000	*****	*****		3.1	3.0	2.9	2.8	2.7	2.6	2.4	1.9	1.1
1,500	*****	*****		2.6	2.5	2.4	2.3	2.2	2.2	2.0	1.5	0.9
2,000	*****	*****		*****	2.2	2.1	2.0	1.9	1.9	1.7	1.3	0.8
3,000	*****	*****		*****	*****	1.7	1.6	1.6	1.5	1.4	1.1	0.6
4,000	*****	*****		*****	*****	*****	*****	1.4	1.3	1.2	0.9	0.5
5,000	*****	*****		*****	*****	*****	*****	*****	1.2	1.1	0.8	0.5
6,000	*****	*****		*****	*****	*****	*****	*****	*****	1.0	0.8	0.4
7,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	0.7	0.4
8,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	0.7	0.4
9,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	*****	0.4
10,000	*****	*****		*****	*****	*****	*****	*****	*****	*****	*****	0.3

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

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

Suppose that the user estimates that 2,920,763 / 8,375,473 = 34.9% of households had a lawn and used chemical fertilizers in 2005. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., households that had a lawn in 2005), it is necessary to use both the percentage 34.9% and the numerator portion of the percentage 2,920,763 in determining the coefficient of variation.

3) The numerator, 2,920,763, 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 3,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 35.0%.

4) The figure at the intersection of the row and column used, namely 1.6% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 1.6%. The finding that 34.9% of households had a lawn and used chemical fertilizers in 2005 can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 2,724,073 / 7,331,151 = 37.2% of households that own their dwelling, had a lawn and used chemical fertilizers in 2005 while 196,690 / 1,044,323 = 18.8% of households that rent, had a lawn and used chemical fertilizers in 2005. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for households that own their dwelling, had a lawn and used chemical fertilizers in 2005 as 1.6%, and the CV of the estimate for households that rent as 6.8%.

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

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

where $\hat{X}_1$ is estimate 1 (households that own their dwelling) $\hat{X}_2$ is estimate 2 (households that rent) 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 $\hat{d} = 0.372 - 0.188 = 0.184$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.372)(0.016)\right]^2 + \left[(0.188)(0.068)\right]^2} \\ &= \sqrt{(0.000035) + (0.000163)} \\ &= 0.014\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.014 / 0.184 = 0.076$

4) So the approximate coefficient of variation of the difference between the estimates is 7.6%. The difference between the estimates is considered acceptable and this estimate can be released with no qualifications.

Example 4: Estimates of Ratios

Suppose that the user estimates that 2,724,073 households that own their dwelling, had a lawn and used chemical fertilizers in 2005, while 196,690 households that rent, had a lawn and used chemical fertilizers in 2005. The user is interested in comparing the estimate of owners versus renters in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of households that own their dwelling, had a lawn and used chemical fertilizers in 2005. The denominator of the estimate ($\hat{X}_2$) is the number of households that rent, had a lawn and used chemical fertilizers in 2005.

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

3) The numerator of this ratio estimate is 2,724,073. The figure closest to it is 3,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 1.6%.

4) The denominator of this ratio estimate is 196,690. The figure closest to it is 200,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 6.8%.

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:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.016)^2 + (0.068)^2} \\ &= \sqrt{0.000256 + 0.004624} \\ &= 0.0698\end{aligned}$$

6) The obtained ratio of households that own their dwelling versus households that rent, had a lawn and used chemical fertilizers in 2005 is 2,724,073 / 196,690 which is 13.8 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 7.0% which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of households that own their dwelling had a lawn and used chemical fertilizers in 2005, to households that rent, had a lawn and used chemical fertilizers in 2005 is 11.0 for Saskatchewan, while it is 13.8 for Manitoba. 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 Saskatchewan ratio ($\hat{R}_1$) and the Manitoba ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Saskatchewan ratio is 17.3% and 24.5% for Manitoba.

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 $\hat{d}$ = 11.0 – 13.8 = -2.8 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(11.0)(0.173)\right]^2 + \left[(13.8)(0.245)\right]^2} \\ &= \sqrt{(3.6214) + (11.4312)} \\ &= 3.880\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 3.880 / (-2.8) = -1.386.

4) So the approximate coefficient of variation of the difference between the estimates is 138.6%. 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.

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 households, that had a lawn and used chemical fertilizers in 2005. (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 34.9% (or expressed as a proportion 0.349)

t = 2

$\alpha_{\hat{x}}$ = 1.6% (0.016 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.349 – (2) (0.349) (0.016), 0.349 + (2) (0.349) (0.016)}

$CI_{\hat{x}}$ = {0.349 – 0.011, 0.349 + 0.011}

$CI_{\hat{x}}$ = {0.338, 0.360}

With 95% confidence it can be said that between 33.8% and 36.0% of households that had a lawn used chemical fertilizers in 2005.

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 households that own their dwelling, had a lawn and used chemical fertilizers in 2005 and the proportion of households that rent, had a lawn and used chemical fertilizers in 2005. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.014. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.372 - 0.188}{0.014} = \frac{0.184}{0.014} = 13.1$$

Since t = 13.1 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the HES 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 a year a household composts would be greater than the coefficient of variation of the corresponding proportion of households that compost their kitchen waste. 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 HES2006_HH_CVTabsE.pdf for the coefficient of variation tables for the household file and to HES2006_PR_CVTabsE.pdf for the person file.

11.0 Weighting

Since the Households and the Environment Survey (HES) 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 where 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 Households and the Environment Survey

The principles behind the calculation of the weights for the HES 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 HES microdata file.

1) An adjustment to account for the use of a sub-sample, instead of the full LFS sample. Depending on the size of the LFS sample of interest in a given domain of interest, different numbers of LFS panels (from 2 to 6) were surveyed for the HES.

2) An adjustment to account for the LFS computer-assisted personal interview (CAPI) cases that were not interviewed for the HES.

3) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the HES for households that did respond to the LFS. All units selected for the HES were modeled using logistic regression to calculate their probability to respond. This probability was used to group records into clusters. The inverse of the observed response rate in each cluster was used as the third adjustment factor.

4) A fourth adjustment is done to match the census projections for independent province-sex-age groups and census metropolitan area (CMA) counts (in a calibration exercise).To produce the final person-weight, a fifth adjustment was made to account for the selection of a single household member for the Transportation Decisions section. Then, the final adjustment used generalized regression (GREG) estimation to calibrate the interim HES person-weights - matching the age-sex distributions for each province and the population counts for several CMAs. These population projections were taken from the same totals used in the LFS. The final HES person-weight is the outcome of these six adjustments to the initial LFS sub-weight.

To produce the final household-weight, the final person-weight is modified by undoing the fifth adjustment above (to return to a household-level for estimation) before a sixth and final adjustment is performed by calibrating to independent estimates of the distribution of households in each region according to size (i.e., one, two, or three or more occupants).

The resulting weight WTHM is the final weight which appears on the HES microdata file for the household estimates. The variable WTPM is the final weight which appears on the person microdata file for estimates relating to the questions on travel to work in the Transportation Decisions section.

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 Households and the Environment Survey Questionnaire

The Households and the Environment Survey (HES) questionnaire was used in February, March and April 2006 to collect the information for the supplementary survey. The file HES2006_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See HES2006_Master_HH_CdBk.pdf for the record layout with univariate counts for the household file and HES2006_Master_PR_CdBk.pdf for the person file.