MICRODATA USER'S GUIDE
THE SURVEY OF HOUSEHOLD ENERGY USE

FEBRUARY 1993

TABLE OF CONTENTS

1. INTRODUCTION

The 1993 Survey of Household Energy Use (SHEU) was conducted by Statistics Canada on behalf of Natural Resources Canada, in cooperation with the provinces of Nova Scotia, New Brunswick, Ontario, Manitoba, Saskatchewan and with SaskPower. This manual is to help with the use of the microdata file of survey results. It is important to become familiar with the manual before releasing any estimates derived from the microdata file of the Survey of Household Energy Use.

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

Elizabeth Majewski
Special Surveys Group
Statistics Canada
5A6 Jean Talon Building
Tunney's Pasture
Ottawa, Ontario, K1A 0T6
Telephone: (613) 951-4584

Yves Bélanger
Special Surveys Group
Statistics Canada
5D6 Jean Talon Building
Tunney's Pasture
Ottawa, Ontario, K1A 0T6
Telephone: (613) 951-9477

André Bourbeau
Energy End Use Analysis and Data Development
Efficiency and Alternative Energy Branch
Natural Resources Canada
15th Floor, 580 Booth Street
Ottawa, Ontario, K1A 0E4
Telephone: (613) 995-7307

2. BACKGROUND

Until the Survey of Household Energy Use was undertaken, Statistics Canada's annual Survey of Household Facilities and Equipment was the only national data source on the stock of energy using equipment and appliances. However, it does not provide sufficient details and does not cover the characteristics of the house needed by Natural Resources Canada for the recently created National Energy Use Database. SHEU was commissioned to enrich the residential sector of this database.

The Survey was administered as a supplement to the February 1993 Labour Force Survey. The target population for this survey was composed of all the housing units in Canada (excluding Yukon and North Western Territories) occupied as primary residences, both owned and rented.

3. SURVEY OBJECTIVES

The objectives of the Survey of Household Energy Use were to provide detailed information about:

. the characteristics of the house that affect air-tightness (thermal envelope),
. space heating equipment and fuels used,
. space cooling equipment,
. water heating equipment,
. stock and features of the major energy consuming appliances,
. use of appliances, heating and cooling.

Additionally, the survey results will be supplemented by two sets of data. The first one will be a file with the average energy consumption of the appliances used by the sampled households that provided make and model information. The second one will be a file containing yearly electricity, natural gas and heating oil consumption of a subset of the surveyed households that pay for their utilities, agreed that Statistics Canada obtains this information from their suppliers, and for whom linking with utility files was successful.

4. CONCEPTS AND DEFINITIONS

Dwelling

A structurally separate set of living premises with a private entrance from outside the building, or from a common hallway or stairway inside.

Type of Dwelling:

Single detached - a house containing one dwelling unit and not attached to any other building or structure.

Double - joined to only one other dwelling (side by side).

Row or terrace - three or more dwellings sharing common walls extending from ground to roof but with no other dwellings either above or below.

Duplex - two dwellings one above the other, not attached to any other structure.

Apartment in a building that has fewer than five storeys - dwellings in triplexes, quadruplexes or apartment buildings.

Apartment in a building that has five or more storeys.

Mobile home - a single dwelling, designed and constructed to be transported on its own chassis.

Household

A person or a group of persons occupying one dwelling unit is defined as a "household". The number of households is equal to the number of occupied dwellings.

Family Status

A family is defined as "a group of two or more persons who are living together in the same dwelling and who are related by blood, marriage (including common-law) or adoption". A person living alone or who is related to no one else in the dwelling where he/she lives is classified as an "unattached individual". The head of a family is determined by the respondent's perception of headship and as such is solely a statistical device which has no economic connotation.

Urban - Rural Area

Statistics Canada defines an urban area as an area which has attained a population concentration of at least 1,000, and a population density of at least 400 per square kilometre, at the previous census. All territory lying outside urban areas is considered rural.

Labour Force Status

Status of the respondent in the labour market : a member of the non-institutional population 15 years and over is designated as either **employed, unemployed** or **not in the labour force.**

Employed

Employed persons are those who, during the reference week:

(a) did any work[1] at all

(b) had a job but were not at work due to:
- own illness or disability
- personal or family responsibilities
- bad weather
- labour dispute
- vacation
- other reason not specified above (excluding persons on layoff and persons whose job attachment was to a job start at a definite date in the future).

---

[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 where unpaid family work is defined as unpaid work which contributed directly to the operation of a farm, business or professional practice owned or operated by a related member of the 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.

Unemployed

Unemployed persons are those who, during the reference week:

(a) were without work, had actively looked for work in the past four weeks (ending with reference week), and were available for work[2];

(b) had not actively looked for work in the past four weeks but had been on layoff[3] and were available for work;

(c) had not actively looked for work in the past four weeks but had a new job to start in four weeks or less from the reference week,and were available for work.

Not in the Labour Force

Those persons in the civilian non-institutional population 15 years of age and over who, during the reference week, were neither employed nor unemployed.

---

[2] Persons in this group meeting the following criteria are regarded as available: (i) were full-time students seeking part-time work who also met condition (ii) below. (Full-time students looking for full-time work are classified as not available for work in the reference week.)

(ii) reported that there was no reason why they could not take a job in reference week, or if they could not take a job it was because of "own illness or disability", "personal or family responsibilities", or "already had a job".

[3] Persons are classified as being on layoff only when they expect to return to the job from which they were laid off.

5. SURVEY METHODOLOGY

The Survey of Household Energy Use was administered in February 1993 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 Sections 5.1 to 5.4[4]. Section 5.5 describes how the Survey of Household Energy Use departed from the basic LFS design in February 1993.

5.1 Population Coverage

The LFS is a monthly household survey whose sample of individuals is representative of the civilian, non-institutionalized population 15 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are residents of the Yukon and Northwest Territories, 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 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 appears within.

---

[4] A detailed description of the LFS design is available in the Statistics Canada publication entitled **Methodology of the Canadian Labour Force Survey,** 1984-1990 (catalogue #71-526).

5.2.1 Primary Stratification

Provinces are first stratified into economic regions - geographic areas of more or less homogeneous economic structure formed on the basis of federal provincial agreements. Economic regions are relatively stable over time.

These economic regions are treated as primary strata and further stratification is carried out within them (see section 5.2.3).

5.2.2 Types of Areas

Economic regions are further disaggregated into 3 categories: self-representing areas (SRU's), non-self-representing areas (NSRU's) and special areas. Generally SRU's are urban areas whose population as of the 1981 Census exceeds 15,000 persons or whose unique labour force characteristics demand their establishment as SRU's. For the most part, SRU boundaries are coincident with delineations established for the Census.

All SRU's in each economic region are included in the survey and, as the name implies, each is represented by its own sample.

NSRU's are the areas lying outside the SRU's and they consist largely of small urban centres and rural areas. Each economic region contains one NSRU which is represented by its own sample.

A small proportion (approximately 1%) of the LFS population is found in institutions (for example, live-in staff of hospitals or schools or permanent residents of hotels or motels), on military bases (civilian personnel only) or in remote areas of provinces which are not readily accessible to LFS interviewers. For administrative purposes, this portion of the population is sampled separately through the special area frame. This portion of the sample is selected on a province-wide basis, without reference to the stratification used for SRU and NSRU areas.

5.2.3 Secondary Stratification

SRU areas are next individually delineated into design strata, which reflect areas of similar socio-economic status as identified in the 1981 Census. The extent of the stratification (i.e. number of strata) depends upon the size of the SRU.

In economic regions in which the NSRU population constitutes a significant proportion of the economic region population, the NSRU is next delineated into separate urban and rural strata. Within each of these strata, further stratification is carried out to reflect differences on a number of labour force characteristics.

In special areas, strata are formed on a province-wide basis. The strata reflect the main types of special groups in the population which require special administrative sampling procedures. These are: military establishments, institutions and remote areas.

5.2.4 Cluster Delineation and Selection

Within each of the secondary strata found in SRU areas, a number of geographic contiguous groups of dwellings, or clusters, are formed based upon a combination of 1981 Census counts and field enumeration. These clusters generally are coincident with city blocks or block faces. The selection of a sample of clusters (generally 6 or 12 clusters) from each of these secondary strata represents the first stage of sampling in SRU areas.

Within each of the secondary strata in NSRU areas, a number of large geographic areas are delineated in such a way that each one reflects the composition of the stratum within which it is located with respect to a number of socio-economic characteristics. Two or four of these areas, known as primary sampling units (or PSU's) are selected into the sample from each secondary stratum. Within each selected PSU, a number of smaller geographically contiguous groups of dwellings, or clusters, are then formed using well-defined physical features which are recognizable both on maps and in the field.

In special areas, census enumeration areas (geographic areas covered by individual enumerators for the Census) represent the first stage of selection. Within those selected, where necessary, geographically contiguous groups of dwellings or clusters are formed and the selection of a sample of these represents the second stage of sampling.

5.2.5 Dwelling Selection

In all three types of areas (SRU, NSRU and special 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 6 dwellings (on average) is then selected. This represents the final stage of sampling.

In the 17 largest SRU's, a sample of apartments in large apartment buildings is selected from a separate register based upon information supplied by CMHC. 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.

5.2.6 Person Selection

Demographic information is obtained for all persons 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.

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 subprovincial level, to meet the requirements of federal, provincial and municipal governments as well as a host of other data users.

The monthly LFS sample consists of approximately 73,000 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 63,000 dwellings remain which are occupied by one or more eligible persons. From these dwellings, LFS information is obtained for approximately 122,000 civilians aged 15 or over.

5.4 Sample Rotation

The LFS employs a panel design whereby the entire monthly sample of dwellings consists of 6 panels, or rotation groups, of approximately equal size. Each of these panels can be considered by itself to be representative of the entire LFS population. All dwellings in a rotation group remain in the LFS sample for 6 consecutive months after which time they are replaced (rotated out of the sample) by a new panel of dwellings selected from the same or similar clusters.

This rotation pattern was adopted to ensure that the sample of dwellings constantly reflects changes in the current housing stock and to minimize any problems of non-response or respondent burden that would occur if households were to remain in the sample for longer than 6 months. It also has the statistical advantage of providing a common sample base for short-term month-to-month comparisons of LFS characteristics.

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 LFS Design for the Supplement

The Survey of Household Energy Use used a subset of the six rotation groups in the February 1993 LFS sample. Depending on the province, from one half of a rotation group to two complete rotation groups were surveyed. In addition, in five provinces, one or two rotation groups that had rotated out of a previous LFS (December 1992 or January 1993) were also used. The following table summarizes the number of rotation groups used in each province.

PROVINCE	Number of rotation groups in the Feb93 LFS	Number of rotation groups in a previous LFS	Total number of rotation groups
Newfoundland	2	0	2
P.E.I.	2	2	4
Nova Scotia	2	1	3
New Brunswick	2	1	3
Quebec	1/2	0	1/2
Ontario	7/10	0	7/10
Manitoba	2	1	3
Saskatchewan	2	1	3
Alberta	1	0	1
B.C.	1	0	1

For the SHEU, the coverage of the LFS was modified to exclude all collective households such as institutions, hotels, etc.

5.6 Sample Size by Province for the Supplement

The following table shows the number of households which were eligible for the SHEU, i.e. they belonged to the target population and to the sampled rotations, and they had responded to the LFS.

PROVINCE	SAMPLE SIZE
Newfoundland	1013
Prince Edward Island	947
Nova Scotia	2165
New Brunswick	1878
Québec	914
Ontario	1987
Manitoba	2065
Saskatchewan	2389
Alberta	913
British Columbia	911
CANADA	15182

Disparities in provincial sample sizes are explained by the fact that several provinces chose to increase their sample size and contributed funding to cover the additional costs. These provinces are Nova Scotia, New Brunswick, Ontario, Manitoba and Saskatchewan.

6. DATA COLLECTION

Data collection for the LFS is carried out each month during the week following the LFS reference week, usually the third week of the month.

6.1 Interviewing for the LFS

Statistics Canada interviewers, who are part-time employees hired and trained specifically to carry out the LFS, contact each of the sampled dwellings to obtain the required labour force information. Each interviewer contacts approximately 70 dwellings per month.

Dwellings new to the sample are contacted through a personal visit. The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all eligible members. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. As a result, approximately 85% of all dwellings are interviewed by telephone. In these subsequent monthly interviews, as they are called, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month.

In all dwellings, information about all household members is obtained from a knowledgeable household member - usually the person at home when the interviewer calls. Such 'proxy' reporting, which accounts for approximately 55% 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.

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.

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.

6.2 Supervision and 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 8 Statistics Canada regional offices.

6.3 Data Collection Modifications for the SHEU

The collection period was extended to two weeks to increase the response rate to the supplement. Approximately a week before the collection, households received in the mail a guide to help them prepare for the interview. As a further step towards obtaining better information, some households were offered a visit of the interviewer instead of the usual interview over the phone. It was assumed that two categories of households would benefit most from interviewer's help: households renting their dwellings where all the household members were at least 65 years old and those where the highest level of education was grade eight.

The SHEU questionnaire was administered to a member of household who, according to the LFS respondent, was most knowledgeable about household equipment, appliances and the house or apartment itself.

6.4 Non-Response to the LFS

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.

6.5 Non-Response to the Survey of Household Energy Use

For households responding to the LFS, the next stage of data collection was to administer the SHEU. In total, 15,182 households were eligible for the supplementary survey; the SHEU interview was completed for 10,982 of these households. This represents a response rate of 72.3%. More detailed information on response rates is presented in Chapter 8 (Data Quality).

7. DATA PROCESSING

The main output of this survey is a "clean" microdata file. This section presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

Capture of survey data was accomplished using minicomputers located in Statistics Canada Head Office in Ottawa. During this process any document containing at least one interviewer-completed item was captured and an unedited version of the computer record was electronically transmitted to the Special Surveys Group for further processing. In total, 15,277 documents were captured and transmitted for the survey[5].

7.2 Editing

The captured data include some interviewer errors and errors made by keyers. Editing corrects these errors to make each record internally consistent.

A "top down" approach to editing was adopted, which means that corrections for missing or improbable values followed the questionnaire flow. The exceptions were appliance and equipment make and model, which were always retained, independently of the answers to the earlier questions.

The questionnaire had separate sections for house residents and for apartment dwellers. There were some records with the wrong sections answered. The editing process moved the answers to the appropriate section.

The edits took also care of conversion of metric to imperial measurements and of recoding to appropriate ranges.

The records that had no or very few answers to the core questions in the questionnaire were dropped.

---

[5] Note that the number of questionnaires captured (15,277) was greater than the number of eligible households (15,182). The difference was made up of households that did not respond to the LFS. All but a few of these households were non-respondents to the SHEU as well, even though a questionnaire was transmitted.

7.3 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. For example, the variable Q85ADV represents the fuel used by the principal heating system. It is derived from variables Q78 (identifying heat pumps), Q84A (identifying electric baseboards and radiant heating) and Q85A (identifying the fuel used by heating equipments other than heat pumps, electric baseboards and radiant heating).

Another example is variable Q183INDR which represents the total number of incandescents light bulbs used indoors. It is simply the sum of variables Q183A to Q183I.

7.4 Weighting

The principle behind estimation in a probability sample such as the LFS is that each dwelling in the sample "represents", besides itself, several other dwellings not in the sample. For example, in a simple random 2% sample of the population, each dwelling in the sample represents 50 dwellings 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 single houses with a refrigerator of large size 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.

7.5 Suppression of Confidential Information

It should be noted that the 'Public Use' microdata files described above differ in a number of important respects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. 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 Section 9 of this document.

Geographic Identifiers

The survey master data file includes explicit geographic identifiers for province, economic region and Census Metropolitan Area. The survey public-use microdata files identify the province and the urban or rural character of the area of residence. For the urban areas the size categories are also provided.

Names of Energy Suppliers

Respondents provided the names of their energy suppliers together with a permission to contact the utilities. This information has been provided for Statistics Canada use only and has been suppressed.

Labour Force Status

"Unemployed" and "Not in the Labour Force" categories have been combined.

Rent

Rent values higher than $999 have been combined in one category.

8. DATA QUALITY

8.1 Response Rates

The following table summarizes the response rates to the Survey of Household Energy Use.

PROVINCE	Number of responding households	Response rate (%)
Newfoundland	837	82.6
Prince Edward Island	732	77.3
Nova Scotia	1518	70.1
New Brunswick	1235	65.8
Québec	720	78.8
Ontario	1407	70.8
Manitoba	1476	71.4
Saskatchewan	1704	71.3
Alberta	674	73.8
British Columbia	679	74.5
CANADA	10982	72.3

Note: The response rate is the number of households that responded to the SHEU as a percentage of the number of households responding to the LFS and that were eligible for the SHEU. The overall response rate to the February 1993 LFS itself was 94.5%.

8.2 Survey Errors

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

Errors which are not related to sampling may occur at almost every phase of a survey

operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

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

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 households who responded to the survey to compensate for those who did not respond.

In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, could not recall the requested information, or could not provide proxy information.

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 78.5% of Canadian homes use a microwave oven, and this estimate is found to have a standard error of .006. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.006}{.785}\right) \; x \; 100\% \; = \; 0.8\%$$

9. PUBLICATION AND RELEASE GUIDELINES

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

9.1 Organization of the Microdata File

The file contains the data from three sources: the LFS, the Rental Survey conducted in conjunction with the LFS, LFS supplement - the SHEU.

The SHEU part of the file reflects the questionnaire: questions 13-77 and 164-195 apply to all the dwellings, while questions 78-142 and 161-163 apply to the dwellings identified as houses, and questions 143-160 apply to dwellings identified as apartments.

To direct a dwelling to the "house" section of the questionnaire or to the "apartment" section, the LFS classification of the type of dwelling and tenure (owned-rented) was used.

The following LFS types of dwellings answered questions for houses:
- single detached,
- double,
- row or terrace,
- **duplex if owned by a household member**,
- mobile home.

The following types of dwellings answered questions for apartments:
- **duplex if rented**,
- low-rise apartment (less than 5 storeys),
- high -rise apartment (5 or more storeys).

Users have to remember that estimates of certain dwelling characteristics require combining data from both the house and the apartment sections. For example, if the proportion of dwellings using electric baseboards as primary heating is to be estimated, one will combine answers from question 84 and question 144.

When, for example, a proportion of duplexes with wood burning fireplaces is to be estimated, one has to combine the counts for question 91 for duplexes only and for question 148, again for duplexes only and to use as a denominator the total number of duplexes.

9.2 Rounding Guidelines

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

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

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

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

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

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

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

9.3 Sample Weighting Guidelines for Tabulation

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

If proper weights are not used, the estimates derived from the microdata tapes 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.

9.3.1 Definitions of types of estimates: Categorical vs. Quantitative

Before discussing how the SHEU 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 SHEU.

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 one-storey houses or the proportion of homes using a microwave oven are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Do you have a wood burning fireplace in your home?
R: Yes / No

Q: How much of the basement area is heated? (Mark one only)
R: Whole basement / More than one half / About one half / Less than half / Don't know

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 units in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of triple pane windows for houses where they are used. The numerator is an estimate of the total number of triple pane windows, and its denominator is the number of houses where triple pane windows are used.

Examples of Quantitative Questions:

Q: How many (wood doors) ... with storm doors?
R: |_|

Q: How many ordinary (incandescent) light bulbs do you have in your ... kitchen?
R: |_|_|

9.3.2 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 X/Y are obtained by:

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

9.3.3 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of triple pane windows used in single houses, multiply the sum of the values reported in Q128A1 (skylights), Q129A1 (oversized) and Q129A2 (other size) – be careful not to count the 9's or 99's corresponding to "not stated"! – by the final weight for the record, then sum this value over all records with DWELLING=1 (single detached houses).

To obtain a weighted average of the form X/Y, the numerator (X) is calculated as for a quantitative estimate and the denominator (Y) is calculated as for a categorical estimate. For example, to estimate the average number of triple pane windows used in single houses,

(a) estimate the total number of triple pane windows as described above,
(b) estimate the number of houses in this category by summing the final weights of all records with DWELLING=1, then
(c) divide estimate (a) by estimate (b).

Note that in this example the average will be calculated over all single detached houses,

whether they actually have triple pane windows or not. A more meaningful estimate may be obtained by restricting the denominator to be the number of single detached houses for which triple pane windows are used, i.e. in (b), summing weights for records with DWELLING=1 and sum of Q128A1, Q129A1 and Q129A2 greater than zero.

9.4 Guidelines for Statistical Analysis

The Survey of Household Energy Use is based upon a complex design, with stratification and 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.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures 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 almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists which can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

In order to provide a means of assessing the quality of tabulated estimates, Statistics Canada has produced a set of Approximate Sampling Variability Tables (commonly referred to as "CV Tables") for the SHEU. These tables can be used to obtain approximate coefficients of variation for categorical-type estimates and proportions. See Chapter 10 for more details.

9.5 CV Release Guidelines

Before releasing and/or publishing any estimate from these microdata tapes, users should first determine the number of respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the **rounded** estimate and follow the guidelines below.

Sampling Variability Guidelines

Type of Estimate	CV (in %)	Guidelines
1. Unqualified	0.0 - 16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Qualified	16.6 - 25.0	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter Q (or in some other similar fashion).
3. Confidential	25.1 - 33.3	Estimates can be considered for general unrestricted release only when sampling variabilities are obtained using an exact variance calculation procedure. Unless exact variances are obtained, such estimates should be deleted and replaced by dashes (---) in statistical tables.
4. Not for Release	33.4 or greater	Estimates cannot be released in any form under any release OR circumstances. In statistical tables, such estimates should be deleted and replaced by dashes(--)

Note: These sampling variability guidelines should be applied to weighted rounded estimates.

10. APPROXIMATE SAMPLING VARIABILITY TABLES

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

The table below shows the design effects, sample sizes and population counts by province which were used to produce the Approximate Sampling Variability Tables.

PROVINCE	DESIGN EFFECT	SAMPLE SIZE	POPULATION
Newfoundland	3.0	837	186070
Prince Edward Island	2.1	732	45736
Nova Scotia	3.0	1518	336080
New Brunswick	2.5	1235	260915
Quebec	2.4	720	2710836
Ontario	2.7	1407	3810478
Manitoba	2.6	1476	402524
Saskatchewan	2.3	1704	368270
Alberta	2.2	674	934816
British Columbia	3.0	679	1303492
Atlantic Provinces	3.0	4322	828801
Prairies	2.6	3854	1705610
Canada	6.0	10982	10359217

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. The use of actual variance estimates would allow users to release otherwise unreleaseable estimates, i.e. estimates with coefficients of variation in the "confidential" range.

Remember: if the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. This is because the formulas used for estimating the variance do not hold true for small sample sizes.

10.1 How to Use the CV Tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the 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 Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the 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 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 single detached houses with a wood burning fireplace is more reliable than the estimated number of single detached houses with a wood burning fireplace. Note that in the tables the cv's 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. a particular type of dwelling), 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 ($\hat{d} = \hat{X}_1 - \hat{X}_2$) is:

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

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 single detached houses and the numerator is the number of single detached houses with a wood burning fireplace.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of houses with a **wood** burning fireplace as compared to the number of houses with a **gas** burning fireplace, the standard deviation 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 R. That is, the standard error of a ratio ($\hat{R} = \hat{X}_1 / \hat{X}_2$) 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 cv's for the two ratios are first determined using Rule 4, and then the cv of their difference is found using Rule 3.

10.2 Examples of Using the CV Tables for Categorical Estimates

The following "real life" examples are included to assist users in applying the foregoing rules.

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

Suppose that a user estimates that 2,321,823 households in Canada use halogen light bulbs (indoors or outdoors). How does the user determine the coefficient of variation of this estimate?

(1) Refer to the cv table for CANADA.

(2) The estimated aggregate (2,321,823) 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 2,000,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 4.8%.

(4) So the approximate coefficient of variation of the estimate is 4.8%. The finding that there were 2,321,823 households in Canada that use halogen light bulbs is publishable with no qualifications.

Example 2 : Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates that 2,849,308 / 5,823,176 = 48.9% of single detached houses in Canada have one storey. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the table for CANADA.

(2) Because the estimate is a percentage which is based on a subset of the total population (i.e., single detached houses), it is necessary to use both the percentage (48.9%) and the numerator portion of the percentage (2,849,308) in determining the coefficient of variation.

(3) The numerator, 2,849,308, does not appear in the left-hand column (the 'Numerator of Percentage' column) so it is necessary to use the figure closet 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 figure closest to it, 50.0%.

(4) The figure at the intersection of the row and column used, namely 3.1% is the coefficient of variation to be used.

(5) So the approximate coefficient of variation of the estimate is 3.1%. The finding that 48.9% of residents of single detached houses have one storey can be published with no qualifications.

Example 3 : Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 3,830,339 / 7,261,564 = 52.7% of houses have their temperature lowered at night, while the same is true for 782,367 / 3,097,653 = 25.3% of apartments. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the CANADA cv table for in the same manner as described in example 2 gives the cv of the estimate for houses as 2.7%, and the cv of the estimate for apartments as 7.5%.

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

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

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.

That is, the standard error of the difference $\hat{d} = (.527 - .253) = .274$ is:

$$\sigma_{\hat{d}} = \sqrt{[(.527)(.027)]^2 + [(.253)(.075)]^2}$$

$$= \sqrt{(.0002025) + (.0003601)}$$

$$= .024$$

(3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}}/\hat{d} = .024 / .274 = .088$.

(4) So the approximate coefficient of variation of the difference between the estimates is 8.8%. This estimate is publishable with no qualifications.

Example 4 : Estimates of Ratios

Suppose that the user estimates that there are 2,053,991 single detached houses with a wood burning fireplace, while there are 316,530 single detached houses with a gas burning fireplace. The user is interested in comparing these two estimates 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 single detached houses with a wood burning fireplace. The denominator of the estimate ($= \hat{X}_2$) is the number of single detached houses with a gas burning fireplace.

(2) Refer to the table for CANADA.

(3) The numerator of this ratio estimate is 2,053,991. The figure closest to it is 2,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 4.8%.

(4) The denominator of this ratio estimate is 316,530. The figure closest to it is 300,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 13.4%.

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

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

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

That is ,

$$\alpha_{\hat{R}} = \sqrt{(.048)^2 + (.134)^2}$$

$$= 0.142$$

The obtained ratio of the number of single detached houses with a wood burning fireplace, over the number of single detached houses with a gas burning fireplace is 2,053,991 / 316,530 which is 6.49:1. The coefficient of variation of this estimate is 14.2%, which is releasable with no qualifications.

10.3 How to Use the CV 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 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as ($\hat{X}$-k, $\hat{X}$+k) 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:

$$CI_X = [\hat{X} - t\hat{X}\alpha_{\hat{X}} , \hat{X} + t\hat{X}\alpha_{\hat{X}}]$$

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 = 3 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.

10.4 Example of Using the CV Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of single detached houses with one storey (from Example 2, section 10.2) would be calculated as follows.

$\hat{X}$ = 48.9% (or expressed as a proportion = .489)

t = 2

$\alpha_{\bar{X}}$ = 3.1% (.031 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\bar{X}}$ = {.489 - (2) (.489) (.031), .489 + (2) (.489) (.031)}

$CI_{\bar{X}}$ = {.489 - .030, .489 + .030}

$CI_{\bar{X}}$ = {.459, .519}

With 95% confidence it can be said that between 45.9% and 51.9% of single detached houses have one storey.

10.5 How to Use the CV 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 $\mathbf{X_1}$ and $\mathbf{X_2}$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\bar{X}_1 - \bar{X}_2$ 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 characteristics are significant.

10.6 Example of Using the CV Tables to Do a T-Test

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of houses where the temperature is lowered at night and the proportion of apartments where the same is done. From example 3, section 10.2, the standard error of the difference between these two estimates was found to be = .024. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{.527 - .253}{.024} = \frac{.274}{.024} = 11.42.$$

Since t = 11.42 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.

10.7 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 Survey of Household Energy Use 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 wood burning fireplaces would be greater than the coefficient of variation of the proportion of households with a wood burning fireplace. Hence if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be 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.

10.8 Release Cut-Off's for the SHEU

The minimum size of the estimate at the provincial, regional and Canada levels are specified in the table below. Estimates smaller than the minimum size given in the "Not Releasable" column may not be released under any circumstances.

Table of Release Cut-offs

PROVINCE	PUBLISHABLE	RELEASABLE WITH QUALIFICATION	NOT RELEASABLE
Newfoundland	21,500	10,000	6,000
Prince Edward Island	4,500	2,000	1,000
Nova Scotia	22,500	10,500	6,000
New Brunswick	18,000	8,000	4,500
Québec	295,500	137,500	79,000
Ontario	251,000	113,500	65,000
Manitoba	24,500	11,000	6,500
Saskatchewan	17,500	8,000	4,500
Alberta	100,000	46,500	26,500
British Columbia	182,000	86,000	50,000
Atlantic Provinces	20,500	9,000	5,000
Prairie Provinces	41,000	18,000	10,500
CANADA	204,000	90,000	51,000

10.9 CV Tables

The approximate sampling variability tables for each province and for Canada are given on the following pages.

11. WEIGHTING

Since the Survey of Household Energy Use used a sub-sample of the 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 LFS

In the LFS, the final weight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the balancing factor for non-response, the rural-urban factor and the province-age-sex 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 the 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 .02 for each person and the records must be weighted by 1/.02=50. 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.

Non-response

Notwithstanding the strict controls of the LFS, some non-response is inevitable, despite all the attempts made by the interviewers. The LFS non-response rate is approximately 5%. For certain types of non-response (eg. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were

actually interviewed. This adjustment is done separately for geographic areas called balancing units. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed. To the extent that this assumption is not true, the estimates will be somewhat biased.

Rural-urban Factor

In NSRUs without sufficient rural and urban population for explicit urban and rural strata to be formed, each primary sampling unit (PSU) is composed of both urban and rural parts. Information concerning the total population in rural and urban areas is available from the 1981 Census for each PSU as well as for each economic region (ER) in which explicit urban/rural stratification is not done. Comparison by ER with the actual 1981 rural or urban census counts indicates whether the selected PSUs over- or under-represent the respective areas. The ratio of actual rural-urban counts is divided by the corresponding estimates. These two factors are computed for each relevant ER at the time of selection of the PSUs and are entered on each sample record according to the appropriate area (rural or urban) of the NSRU. Changes in these factors are incorporated at the time of PSU rotations.

LFS 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.

Subprovincial 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. In particular, estimates are produced of the total number of persons aged 15 and over in provincial economic regions and the 24 large metropolitan areas as well as of designated age-sex groups in each of the ten provinces.

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 linear regression model. The regression is set up to ensure that the final weights it produces sum to the census projections for the auxiliary variables, namely various age-sex groups, economic regions and census metropolitan areas.

This weighting procedure ensures consistency with external Census counts, and also ensures that every member of the economic family is assigned the same weight.

11.2 Weighting Procedures for the SHEU

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

(1) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the SHEU for households which did respond to the LFS or for which previous month's LFS data was brought forward.

(2) A readjustment to account for independent projections by type of dwelling at the province level, after the above adjustment has been made. This readjustment also implicitly corrects for the fact that a sub-sample of the full LFS sample was used.

Adjustment (1) is taken into account by multiplying the LFS sub-weight for each responding SHEU record by:

$$\frac{\textit{sum of LFS subweights from each household eligible to the SHEU}}{\textit{sum of LFS subweights from each household responding to the SHEU}}$$

to obtain a non-response adjusted SHEU sub-weight (WEIGHT1).

Adjustment (2) is calculated by multiplying WEIGHT1 for each SHEU responding household by :

$$\frac{\textit{number of dwellings for province-type } i}{\textit{sum of WEIGHT1 for responding dwellings in province-type } i}$$

In practice, types of dwelling were combined as follows for adjustment (2): single detached (type=1), apartments (type=5 or 6) and other (type=2,3,4 or 9), except in Newfoundland where only two categories were used: single detached and other (including apartment).

The resulting weight is the final weight which appears on the SHEU microdata file.

12. RECORD LAYOUT AND UNIVARIATES

13. QUESTIONNAIRES

1) Household Record Docket (Form 03) and Code Sheet
2) The Labour Force Survey Questionnaire (Form 05) and Code Sheet
3) Rent Questionnaire (Form 04)
4) Survey of Household Energy Use, 1993 Questionnaire (Form F08)