Table of Contents

1.0 Introduction

The Survey of Work Arrangements was conducted by Statistics Canada in November 1995 with the cooperation and support of Human Resources Development Canada. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

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

Human Resources Development Canada
Brenda Lipsett
Applied Research Branch
140 promenade du Portage
Phase IV, 4th Floor
Hull, Québec K1A 0J9
(819) 994-4454

2.0 Background

The need for information on work arrangements such as work schedules, flexitime and home-based work was behind the 1991 Survey of Work Arrangements - the first national survey covering these issues. An interest in changes in work arrangements, as well as a need for data on other aspects of working conditions led to the 1995 Survey of Work Arrangements. Both surveys were conducted as supplements to the Labour Force Survey (LFS).

3.0 Objectives

For paid workers, the 1995 Survey of Work Arrangements provides information on issues also covered by the 1991 survey such as:

- when people work (days of the week, hours of work);
- how much control they have over their schedules (e.g. “on call”, flexible schedule);
- who usually works some or all the time at home, and why;
- how many people work paid overtime and how they are compensated for it;
- who has a permanent job and who has a temporary one; and
- how many people hold down two jobs and why they do so.

Additionally, the 1995 survey gathered information not collected by the previous survey on: firm size, employee benefits, unpaid overtime, and preference for fewer or more hours of work.

In order to allow for comparability with the redesigned LFS (to be fully implemented by January 1997) the 1995 SWA used, whenever applicable, the new LFS questions. Consequently, in a few instances this has somewhat limited comparability with the 1991 SWA.

For the self-employed who were not part of the 1991 survey, the 1995 SWA provides information on the days of the week they work, number of employees the business had in the reference week, home-based work, and reasons for self-employment.

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 are described in Section 4.1 while those specific to the Survey of Work Arrangements are given in Section 4.2. Users are referred to Chapter 12 of this document for a copy of the actual survey forms used.

4.1 Labour Force Survey Concepts and Definitions

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 starting 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.

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 1984, these statistics have been based on the 1980 Standard Occupational Classification and the 1980 Standard Industrial Classification. Prior to 1984, the 1971 Standard Occupational Classification and the 1970 Standard Industrial Classification were used.

Reference Week

Entire calendar week 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.

---

[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.

Full-time

Full-time employment consists of persons who usually work 30 hours or more per week at their main job or sole job. Note the difference: in 1991 the designation of full-time employment was applied to all persons who usually worked 30 hours or more a week **at all jobs**, and those who considered themselves to be full-time workers even though their total hours were usually less than 30 per week.

Part-time

Part-time employment consists of all other persons who usually work less than 30 hours per week at their main or sole job.

4.2 SWA Concepts and Definitions

Work Schedules

Regular daytime

Work begins in the morning and ends in the afternoon; the standard 9 to 5 schedule is included in this category.

Regular evening shift

Work starts about 3 or 4 pm and is over by midnight.

Regular night or graveyard shift

Work starts around midnight and finishes around 8 am.

Rotating shifts

A combination of the above shifts provided the shifts rotate on a regular basis and one shift does not predominate over the other(s).

Split shift

Two or more distinct periods of work with a period of free time that is not solely a lunch break, between work periods.

On call

Hours vary substantially from one week to the next. Workers are asked to work as the need arises, not on a prearranged schedule.

Irregular schedules

No regular schedule but a schedule usually arranged one week or more in advance.

Hours Worked

Respondents were instructed to include breaks but to exclude lunch.

Flexible Schedule

A flexible schedule allows workers to choose their starting and stopping times within limits established by the management.

Job Sharing Arrangement

Job sharing implies a voluntary arrangement whereby two or more employees agree to share the job hours of one job.

Job sharing should not be confused with work sharing in which all workers work fewer hours to avoid layoffs.

5.0 Survey Methodology

The Survey of Work Arrangements (SWA) was administered in November, 1995 to a sub-sample of the dwellings in the Labour Force Survey 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]. Sections 5.5 and 5.6 describe how the SWA departed from the basic LFS design in November 1995.

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[5] 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 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.

---

4 A detailed description of the previous LFS design is available in the Statistics Canada publication entitled **Methodology of the Canadian Labour Force Survey,** 1984-1990 (Catalogue no. 71-526-XPB).

5 Since 1992, the LFS has been administered in the Yukon, using an alternative methodology that accommodates some of the operational difficulties inherent to remote locales. To improve reliability due to small sample size, estimates are available on a three-month average basis only. These estimates are not included in national totals.

5.2.1 Primary Stratification

Provinces are divided into economic regions and employment insurance regions. Economic regions (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. Employment insurance economic regions (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 EIER regions for the use of Human Resources 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 (CMAs), 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 divided into two types: those using an apartment list frame and an area frame, as well as those using only an area frame.

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 which is based upon information supplied

by CMHC and is maintained in 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 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 proportionally 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(EAs) 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 six 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 ten 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 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. Response burden is minimized for the elderly (70 years of age or older) by carrying forward their responses from 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 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 59,000 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 52,350 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 employs a panel design whereby the entire monthly sample of dwellings consists of six panels, or rotation groups, of approximately equal size. Each of these panels is, by itself, representative of the entire LFS population. All dwellings in a rotation group remain in the LFS sample for six 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 minimize any problems of non-response or respondent burden that would occur if households were to remain in the sample for longer than six months. It also has the statistical advantage of providing a common sample base for short-term month-to-month comparisons of LFS characteristics, since five of the six rotation groups in the LFS sample are common from month to month.

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 Work Arrangements used three of the six rotation groups in the November 1995 LFS sample. For the SWA, the coverage of the LFS was modified to include all eligible members of the household 15 to 69 years of age. The respondents to the survey were civilian from three selected rotations who were 15 to 69 years of age and who were either paid workers or self employed in their main job.

5.6 Sample Size by Province for the Supplement

The following table shows the number of household members in the LFS sampled rotations who were eligible for the Survey of Work Arrangements.

Province	Sample size
Newfoundland	840
Prince Edward Island	833
Nova Scotia	1,703
New Brunswick	1,635
Quebec	5,310
Ontario	9,150
Manitoba	2,131
Saskatchewan	1,820
Alberta	2,518
British Columbia	2,667
CANADA	**28,607**

There are 42,324 records on the SWA file. Of these, 25,721 records have SWA information (for 3,336 respondents eligible for SWA we did not obtain or did not keep the data). The remaining records are for respondents who were not eligible for the SWA, but who were members of households with SWA respondents and were 15 to 69 years old. All the SWA variables on those records contain 6's - “valid skip” codes.

To select SWA records use the following variables:

LFSSTAT=1 (employed), LFSACTIV= 1 or 2 (worked at a job or was absent from work in the reference week) and COWMAIN= 1-6 (paid worker or self-employed in the main job).

6.0 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 eight Statistics Canada regional offices.

6.3 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.

6.4 Data Collection Modifications for the Survey of Work Arrangements

For households responding to the LFS, the next stage of data collection was to administer the SWA to all eligible household members. An eligible household member was one who was a civilian, 15 to 69 years of age and who was classified as a paid worker or self-employed in their main job.

Interviews were conducted using computer-assisted interviewing. If the selected person was not available, proxy interviews were allowed.

6.5 Non-response to the Survey of Work Arrangements

In total, 28,607 individuals were eligible for the supplementary survey; the SWA information is available for 25,721 individuals. This represents a response rate of 89.9%. More detailed information on response rates is presented in Chapter 8 (Data Quality).

7.0 Data Processing

The main output of the Survey of Work Arrangements is a “clean” microdata file. This section presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

The survey responses were entered during the computer-assisted interviewing. The data capture program automatically followed the flow of the SWA questionnaire and allowed for checking if the codes entered were within a valid range. Additionally, when the discrepancy between the daily number of hours worked (question 15A) and the difference between usual start and stop times (questions 16A and 17A) was over two hours, the interviewers had to correct the entries or to provide explanation. Interviewers transmitted the data from their machines to the regional offices of Statistics Canada, and next, to head office.

7.2 Editing

The first stage of survey processing undertaken at head office was the link with the Labour Force Survey edited file. Records that did not match the final LFS file had to be dropped as well as records that did not have responses to the SWA.

Next, there was a series of steps reformatting complex questions and recoding all the SWA questions according to the following scheme:

0-4 valid codes;
5- not stated;
6- valid skip (question not applicable);
7- don’t know;
8- refusal; and
9- non-response due to a previous “don’t know” or “refusal”, for the LFS variables: not applicable.

Further editing included conversion to the 24 hour clock in those cases when interviewers failed to use it. Wage and salary values, when very low or very high, have undergone case by case check with reference to occupation of respondent . Evident errors were corrected or the values were changed to “not stated”.

The editing of the SWA did not involve imputation for missing values.

7.3 Coding of Open-ended Questions

There were no open-ended questions in the SWA questionnaire. However, there were several questions with “other, specify” answer category. These answers were examined and recoded into additional categories. Such recoding was done for question 21 (reasons for working at home), question 32 (way in which the job is not permanent), question 48 (reason for self-employment), and in question 50 (reason for more than one job). If needed, the added categories may be regrouped as “other” .

7.4 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.

“DVstart” is an example of a simple type of derived variable created by grouping start times into 11 categories.

“SWAQ1145” combines answers from two questions and indicates the number of days worked per week.

“Normal” (normal schedule) combines answers from several questions:

- question 11(works Monday to Friday),
- question 13 (regular daily schedule), questions 16 and 17 (usually begins and ends work at the same time),
- question 16A (starts work between 6:45 and 9:15), and LFS weekly hours (30 to 49).

“Custom1” (federal jurisdiction) combines SIC codes:

- transportation: 451 to 456, 461, 471;
- communication: 481, 482, 484;
- banking: 701 to 709.

Hourly wage has been derived for all those respondents who are not paid by the hour. For example, it has been derived from an annual pay by dividing it by 52 (weeks), and by the usual weekly hours as reported in the LFS interview. Similarly, the weekly wage has been derived by multiplying derived hourly wage by weekly hours and the yearly wage by multiplying weekly wage by 52.

7.5 Weighting

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

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

Details of the method used to calculate these weights are presented in Chapter 11.

7.6 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 Chapter 9 of this document.

The record layout contains an indication “suppressed” for each variable not available on the public use microdata file, but present on the “master” file. For example, the record layout includes explicit geographic identifiers for province, economic region and Census Metropolitan Area. The economic region and the CMA are suppressed on the public access microdata file.

In several instances, the public access microdata file contains only grouped values while the “master” file has the original values. For example, the three digit “industry code” has been suppressed, but the microdata file provides the “industry codes” grouped into 16, 30 and 52 categories.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey and to the Survey of Work Arrangements.

	Household response rate for full LFS (Nov. 95) (*1)	Household response rate for LFS rotations (1,2 & 3 (*1)	Household response rate to SWA (*2)	Number of eligible respondents in SWA	Eligible person response rate to SWA (*3)
Newfoundland	97.8%	98.4%	95.1%	767	91.3%
Prince Edward Island	98.2%	98.3%	94.2%	764	91.7%
Nova Scotia	96.5%	96.8%	95.7%	1,600	93.9%
New Brunswick	96.6%	97.0%	93.2%	1,484	90.8%
Quebec	96.0%	96.7%	93.1%	4,824	90.8%
Ontario	96.1%	97.4%	91.4%	8,090	88.4%
Manitoba	97.3%	98.1%	90.9%	1,914	89.8%
Saskatchewan	97.6%	98.3%	93.4%	1,665	91.5%
Alberta	97.7%	98.7%	93.0%	2,310	91.7%
British Columbia	95.6%	96.8%	90.0%	2,303	86.3%
CANADA	**96.5%**	**97.4%**	**92.4%**	**25,721**	**89.9%**

Note:

(*1) Response rate is number of responding households as a percentage of number of eligible households.

(*2) Response rate is number of households responding to SWA as a percentage of number of households responding to LFS in rotations sampled.

(*3) Response rate is number of eligible individuals responding to SWA as a percentage of number of eligible individuals responding to LFS in rotations sampled, where “eligible” for SWA means being a paid

worker or self employed in the main job, age 15 to 69.

8.2 Survey Errors

The estimates derived from this survey are based on a sample of households. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The difference between the estimates obtained from the sample and the results 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.

8.2.1 Sampling Errors

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 (C.V.) 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 4.2% of Canadians with paid-worker jobs worked regular evening shifts for their main job, and this estimate is found to have standard error of .003. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.003}{.042}\right) \; x \; 100\% \; = \; 7.14\%$$

8.2.2 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.

8.2.3 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response.

Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of 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, or could not recall the requested information.

9.0 Guidelines for Tabulation, Analysis and Release

This section 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 file. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

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

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

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

(c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or

denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the SWA 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 file 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.

9.2.1 Definitions of Types of Estimates: Categorical vs. Quantitative

Before discussing how the SWA 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 Survey of Work Arrangements.

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 paid workers who work regular evening shifts for their main job or the proportion of paid workers working a particular type of schedule 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: Within established limits, can ... choose the time he/she begins and ends his/her work day?
R: Yes / No

Q: What is the main reason that ... works this schedule?
R: Earn more money
Care for children
Care for other family members
Allow time for school
Requirements of the job / no choice
Other reasons

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 days per week usually worked at the person's main job. An example of a quantitative estimate of a ratio is the average hourly wage for paid workers working evening shifts; in this case, the numerator (X) is an estimate of the total of hourly wages for paid workers working evening shifts and the denominator (Y) is an estimate of the number of paid workers working evening shifts.

Examples of Quantitative Questions:

Q: Excluding tips and commissions, what is ....'s hourly rate of pay?
R: $...,..

Q: Last week, how many hours of paid overtime did ... work at this job?
R: |_|_| hrs |_|_| min

9.2.2 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.

9.2.3 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 hours of paid overtime usually worked per week by those who usually work paid overtime, multiply the value reported in SWA-Q24D (hours per week) by the final weight for the record, then sum this value over all records with SWAQ24C=1 (usually works paid overtime).

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 hours of paid overtime worked per week in main jobs,

(a) estimate the total number of hours as described above,
(b) estimate the number of people in this category by summing the final weights of all records with Q24C=1, then
(c) divide estimate (a) by estimate (b).

9.3 Guidelines for Statistical Analysis

The Survey of Work Arrangements 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 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. Variances for simple estimates such as totals, proportions and ratios (for qualitative variables) are provided in the accompanying Sampling Variability Tables.

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

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

- select all respondents from the file who reported SEX=male,
- calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported SEX=male,
- for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight,
- perform the analysis for these respondents using the RESCALED weight.

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

The calculation of truly meaningful 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.

9.4 C.V. Release Guidelines

Before releasing and/or publishing any estimate from the Survey of Work Arrangements, 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. 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 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 weighted rounded 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 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 16.6% to 33.3%. Estimates should be flagged with the letter M (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 U (or some similar identifier) and the following warning should accompany the estimates: “The user is advised that . . . (specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable, and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data.”

10.0 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 (C.V.) 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	1.93	1,855	416,198

Prince Edward Island	1.60	1,226	93,882
Nova Scotia	1.55	2,859	653,524
New Brunswick	1.61	2,680	536,474
Quebec	2.13	8,518	5,292,428
Ontario	2.19	12,743	7,919,751
Manitoba	2.12	2,852	749,456
Saskatchewan	1.80	2,544	653,095
Alberta	1.67	3,316	1,915,045
British Columbia	1.61	3,731	2,667,017
Atlantic provinces	1.65	8,620	1,700,078
Prairies	1.97	8,712	3,317,596
CANADA	**2.25**	**42,324**	**20,896,870**

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 C.V. 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 “paid workers who work a regular evening shift at their main job” is more reliable than the estimated number of “paid workers who work regular evening shift at their main job”. (Note that in the tables the C.V.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 C.V. of the proportion or

percentage is the same as the C.V. of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

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

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference ($\hat{d} = \hat{X}_1 - \hat{X}_2$) 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 paid workers and the numerator is the number of "paid workers who work a regular evening shift at their main job".

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of "full-time paid workers who work a regular evening shift at their main job" as compared to the number of "part-time paid workers who work a regular evening shift at their main job", 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 C.V.s for the two ratios are first determined using Rule 4, and then the C.V. of their difference is found using Rule 3.

10.1.1 Examples of Using the C.V. Tables for Categorical Estimates

The following 'real life' examples are included to assist users in applying the foregoing rules. A public or private employee will be referred to as paid worker in the examples given below.

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

Suppose that a user estimates that 560,099 paid workers work a regular evening shift at their main job. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the C.V. table for CANADA.

(2) The estimated aggregate 560,099 does not appear in the left-hand column (the 'Numerator of Percentage' column), so it is necessary to use the figure closest to it, namely 500,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 4.6%.

(4) So the approximate coefficient of variation of the estimate is 4.6%.

The finding that there were 560,099 paid workers who work a regular evening shift at their main job is publishable with no qualifications.

Example 2 : Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates that (311,385/560,099)=55.6% of paid workers who work a regular evening shift at their main job are employed full time. 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., paid workers

who work a regular evening shift at their main job), it is necessary to use both the percentage 55.6% and the numerator portion of the percentage 311,385 in determining the coefficient of variation.

(3) The numerator, 311,385 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 300,000. 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 4.3% is the coefficient of variation to be used.

(5) So the approximate coefficient of variation of the estimate is 4.3%. The finding that 55.6% of paid workers who work on a regular evening shift at their main job are employed full time can be published with no qualifications.

Example 3 : Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 311,385/560,099=55.6% of paid workers who work a regular evening shift at their main job are employed full time, while 166,970/204,718=81.5% of paid workers who work a regular night shift at their main job are employed full time. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the CANADA C.V. table in the same manner as described in Example 2 gives the C.V. of the estimate for evening shift work as 4.3%, and the C.V. of the estimate for night shift work as 2.7%.

(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

$\hat{d} = (.815-.556)$
$= .259$

is:

$$\sigma_{\hat{d}} = \sqrt{[(.556)(.043)]^2 + [(.815)(.027)]^2}$$

$$= \sqrt{(.000571) + (.000484)}$$

$$= .032$$

(3) The coefficient of variation of $\hat{d}$ is given by

$\sigma_{\hat{d}}/\hat{d} = .032 / .259$
$= 0.123$

(4) So the approximate coefficient of variation of the difference between the estimates is 12.3%. This estimate can be published with no qualifications.

Example 4 : Estimates of Ratios

Suppose that the user estimates that 6,666,982 paid workers work full time on a regular daytime shift at their main job, while 867,674 paid workers work part time on a regular daytime shift at their main job. The user is interested in comparing the estimate of full time versus part time workers on daytime shifts 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 paid workers that work full time on a regular daytime shift at their main job. The denominator of the estimate (= $\hat{X}_2$) is the number of paid workers who work part time on a regular daytime shift at their main job.

(2) Refer to the table for CANADA.

(3) The numerator of this ratio estimate is 6,666,982. The figure closest to it is 7,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 1%.

(4) The denominator of this ratio estimate is 867,674. The figure closest to it is 1,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 3.2%.

(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 ,

$$\alpha_{\hat{R}} = \sqrt{(.010)^2 + (.032)^2}$$

$$= 0.033$$

The obtained ratio of full time versus part time paid workers on a regular daytime schedule at their main job is 6,666,982/867,674 which is 7.68 : 1. The coefficient of variation of this estimate is 3.3%, which is releasable with no qualifications.

10.2 How to Use the C.V. 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.

10.2.1 Example of Using the C.V. Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of paid workers who work full time on a regular evening shift at their main job (from Example 2, Section 10.1.1 would be calculated as follows:

$\hat{X}$= 55.6% or expressed as a proportion = .556

t = 2

$\alpha_{\hat{X}}$ = 4.3% (.043 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

CI_X = {.556 - (2) (.556) (.043), .556 + (2) (.556) (.043)}

CI_X = {.556 - .048, .556 + .048}

CI_X = {.508, .604}

With 95% confidence it can be said that between 50.8% and 60.4% of paid workers on a regular evening shift at their main job work full time.

10.3 How to Use the C.V. 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 $\hat{X}_1 - \hat{X}_2$ 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 characteristics are significant.

10.3.1 Example of Using the C.V. 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 full time paid workers on an evening shift and the proportion of full time paid workers on a night shift. From Example 3, Section 10.1.1 the standard error of the difference between these two estimates was found to be .032, Hence

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{.556 - .815}{.032} = \frac{-.259}{.032} = -8.09.$$

Since t = -8.09 is less than -2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the SWA 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 weeks absent from work would be greater than the coefficient of variation of the corresponding proportion of paid workers with an absence. 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.

10.5 Release Cut-offs for the Survey of Work Arrangements

The minimum size of the estimate at the provincial, regional and Canada levels are specified in the following table. 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	Unqualified	Qualified	Confidential	Not releasable
Newfoundland	15,500 +	7,000 - 15,499	4,000 - 6,999	under 4,000
Prince Edward Island	4,500 +	2,000 - 4,499	1,000 - 1,999	under 1,000
Nova Scotia	13,000 +	5,500 - 12,999	3,000 - 5,499	under 3,000
New Brunswick	11,500 +	5,000 - 11,499	3,000 - 4,999	under 3,000
Quebec	48,000 +	21,000 - 47,999	12,000 - 20,999	under 12,000
Ontario	49,500 +	21,500 - 49,499	12,500 - 21,499	under 12,500
Manitoba	20,000 +	9,000 - 19,999	5,000 - 8,999	under 5,000
Saskatchewan	16,500 +	7,500 - 16,499	4,000 - 7,499	under 4,000
Alberta	35,000 +	15,500 - 34,999	8,500 - 15,499	under 8,500
British Columbia	41,500 +	18,500 - 41,499	10,500 - 18,499	under 10,500
Atlantic provinces	12,000 +	5,000 - 11,999	3,000 - 4,999	under 3,000
Prairie provinces	27,500 +	12,000 - 27,499	7,000 - 11,999	under 7,000
CANADA	**40,500 +**	**18,000 - 40,499**	**10,000 - 17,999**	**under 10,000**

10.6 C.V. Tables

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

11.0 Weighting

Since the Survey of Work Arrangements 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, 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 (e.g. household temporarily absent, refusal), data from a previous month's

interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas, which are defined by employment insurance region, type of area, and the 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. To the extent that this assumption is not true, the estimates will be somewhat biased.

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 15+ 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 calibration method. This method ensures 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, and rotation groups.

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

11.2 Weighting Procedures for the Survey of Work Arrangements

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

(1) An adjustment to account for the use of a three-sixth sub-sample, instead of the full LFS sample.

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

Adjustments (1) and (2) are taken into account by multiplying the LFS sub-weight for each responding SWA record by:

$$\frac{\textit{sum of LFS subweights of eligible persons responding to the LFS}}{\textit{sum of LFS subweights of eligible persons responding to the SWA}}$$

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

At this stage the weight is comprised of two components: the inverse of the sampling rate and the non-response adjustment. A third component, the family weighting adjustment described below was added to improve accuracy of estimates.

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. Using a linear regression model auxiliary information is used to arrive at the final weight. 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. For the SWA, an additional auxiliary variable, number of employed persons by province was used. This final adjustment to the weights improves the reliability of estimates that can be produced by the SWA. At the same time as ensuring consistency with external Census counts, the family weighting procedure also ensures that every member of the household is assigned the same weight.

The resulting weight (FINWT) is the final weight which appears on the Survey of Work Arrangements microdata file.

12.0 Questionnaires and Code Sheets

- Household Record Docket (Form 03)
- The Labour Force Questionnaire (Form 05)
- Supplementary Questionnaire (Form 06)

12.1 The Household Record Docket (Form 03)

The Household Record Docket (Form 03) is used to list all household members whose usual place of residence is the selected dwelling. It is both a survey operations control document and a record of socio-demographic information on household members.

12.2 The Labour Force Questionnaire (Form 05)

The Labour Force Questionnaire (Form 05) is used to collect information on the current and most recent labour market activity of all household members 15 years of age or older. The Form 05 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.

12.3 The Supplementary Survey Questionnaire

The Survey of Work Arrangements questionnaire was used in November 1995 to collect the information for the supplementary survey.

13.0
Record Layout and Univariates