Table of Contents

1.0 Introduction

The Follow-up of 1990 Graduates (F90G) Survey was conducted by the Special Surveys Division, Statistics Canada, on behalf of Human Resources Development Canada (HRDC). It was conducted under the voluntary provisions of the Statistics Act, Revised Statutes of Canada, 1985, Chapter S19. Collection plans for the survey are registered under collection registration number STC/SSD-040-75034 and personal information bank number STC/P-PU-100.

This documentation manual contains information to access and manipulate data from the survey. Anyone interested in obtaining more information may contact the following persons:

Bill Magnus
Special Surveys Division
Statistics Canada
5D6 Jean Talon Building
Tunney's Pasture
K1A 0T6
Tel: (613) 951-4577
Fax: (613) 951-0562
E-mail: magnwil@statcan.ca

Patrice de Broucker
Centre for Education Statistics
Statistics Canada
17th Floor, R.H.Coats Building
Tunney's Pasture
Ottawa, Ontario
K1A 0T6
Tel: (613) 951-3999
Fax: (613) 951-9040
E-mail: debrpat@statcan.ca

Phil Jennings
Human Resources Development Canada
140 Promenade du Portage
5th Floor, Place du Portage, Phase IV
Hull, Quebec
K1A 0J9
Tel: (819) 994-4473
Fax: (819) 953-8584
E-mail: Philip.Jennings@spg.org

2.0 Background

In 1978, Statistics Canada conducted a survey on the labour market experiences of 1976 graduates from universities and community colleges in Canada. In 1984, a similar survey, the National Graduates Survey (NGS) of 1982 graduates was sponsored jointly by the Department of the Secretary of State and Employment and Immigration Canada and conducted by Statistics Canada. The 1984 NGS expanded on the content of the previous survey and extended the population base to include completers of trade/vocational programs in addition to graduates from community colleges and universities.

Since these two surveys in 1978 and 1984, a series of graduate surveys have been completed on the labour market experiences of graduates from universities and community colleges in Canada. The following is a summary of the graduate surveys conducted by Statistics Canada for HRDC.

GRADUATION YEAR	SURVEY YEAR	SURVEY NAME
1976	1978	Survey of 1976 Graduates of Post-Secondary Programs
1982	1984	Survey of 1982 Graduates (also known as the National Graduates Survey or NGS)
1982	1987	Follow-up of 1982 Graduates (FOG87)
1986	1988	Survey of 1986 Graduates (S86G)
1986	1991	Follow-up of 1986 Graduates (F86G)
1990	1992	Survey of 1990 Graduates (S90G)
1990	1995	Follow-up of 1990 Graduates (F90G)

The purpose of all these surveys is to provide information on the integration of recent graduates or completers into the labour market, in terms of the match between education or training and occupation. The need for such information is particularly acute due to continuous changes in labour market conditions, technology and hence, the demand for highly skilled or qualified labour in Canada.

The surveys have provided a wealth of current labour market information in terms of new supplies of highly skilled or qualified labour. They have provided the basis for analyses relating to employment and under-employment conditions for males and females; transitions from fields of study to occupations, comparisons of entry level conditions among various occupations, etc. The data have also been used in order to estimate the parameters of a model projecting flows of students into the labour market. The results of the surveys are used at all levels of government and within the educational community to develop effective educational policies to meet the training needs of all Canadians.

HRDC set aside some resources for "Data and Monitoring" activities in order to provide the information and analysis necessary to support educational and training policies and programs relevant to the social and economic realities of the labour market. Furthermore, provincial governments, other federal departments and other individuals and organisations expressed interest in obtaining relevant data on the transition of Canadians from the postsecondary education system into the labour market.

The Follow-up of 1990 Graduates survey, conducted in May-June 1995, updated the information obtained in the 1992 survey, covering the period between June 1992 and May 1995. It greatly extends the scope of analysis and provides data necessary to carry out:

- cross-sectional analyses five years after graduation;

- aggregate comparisons of short- vs. longer-term experiences of 1990 graduates (comparisons of the 1992 survey results with those of the Follow-up);

- longitudinal analyses of individuals' labour market related experiences (e.g. occupational transitions; geographic mobility; and the evolution of attitudes and satisfaction).

3.0 Objectives

1. To obtain information for longitudinal analysis of a critical labour market group at a key time, focusing on employment, occupation and geographic shifts. These data and analyses will be useful for policy development with respect to education/training and the labour market.

2. To obtain information on the relationship between education and training and labour market experiences and the exposure of graduates to additional training in general.

3. To extend available information required to develop occupational supply and demand projection models and to conduct related studies of supply-demand imbalances in the labour market.

4. To obtain data regarding longer-term labour-market experiences of graduates, with special emphasis on employment and occupations, for use in counselling on careers and postsecondary education course selection.

5. To obtain information on labour-market experiences of members of designated groups (such as women, aboriginal peoples, persons with disabilities and persons in a visible minority), which permits longitudinal and comparative analyses useful in the formulation of job equity policies.

4.0 Concepts and Definitions

This chapter outlines the content areas of the questions for the Follow-up of 1990 Graduates. Users are referred to Chapter 12 of this document for a copy of the actual survey questions.

4.1 Content

The Follow-up of 1990 Graduates survey questionnaire was made up of nine content components or sections of questions. The first seven sections (A-G) consisted of an in-depth exploration of labour market experiences of respondents, whereas the last two sections (H-I) covered information related to education and general topics.

In order to ensure that questions were asked in a manner relevant to respondents' particular circumstances, the questionnaire, in several instances, contained more than one version of the same group of questions. Any one respondent was asked only one of those versions. Therefore, no respondent answered all questions. For example, respondents who have worked continuously for the same employer from June 1992 to May 1995, were, with a few exceptions, requested to respond to questions in Sections A, B, G, H and I. All others were channelled through Section A, C, D, E, F, G, H and I. Section B obtained much the same information as the combined Sections C to F. Both sets established labour market activities during the period and solicited some qualitative information regarding links between fields of study and occupation.

In terms of content, the important difference between the group channelled through Section B and the group channelled through Sections C to F is that the former consisted of individuals whose major labour market activity was already known (by virtue of their continuous employment with the same employer), whereas for the other group, unknown changes would have occurred during the reference period. Thus Sections C to F solicited more specific details in order to obtain a sufficient understanding of respondents' labour market experiences over the period.

The following are short summaries of the contents of each of the eight sections.

Section A: Last Week
Respondents are asked questions about their labour force activity during the week before interview week. Included are questions on full-time work, layoff, future job starts and job search activity.

Section B: Same Job as in June 1992
Respondents who have worked continuously for the same employer since June 1992 answer questions from this section. Where applicable, they are asked for details of their new occupation for the same employer (i.e., work description, start date). Most respondents are asked questions about related work experience and education requirements for their job, job satisfaction and remuneration. Additional questions are asked of those whose job involves periods of part-time and full-time work.

Section C: Details of Current Job
Generally, respondents who have changed employers or had breaks in employment since the June 1992 interview answer questions from this section. These relate to the description of their job, work experience and education requirements for the job, job satisfaction and remuneration.

Sections D, E, F: Other Job Information
Respondents are asked questions about additional jobs held since 1992. Also included are questions regarding the description of these jobs, job earnings and number of hours worked per week.

Section G: Activities in 1994
Respondents are asked questions on their labour force activity during the calendar year 1994. Included are questions on periods of labour force inactivity (i.e., without a job and not looking for one), periods when respondents may have attended school full-time and periods when they were waiting to start a job.

Section H: Education Taken Since June 1992
All respondents are asked questions from this section. These relate to additional education qualifications obtained since June 1992, the importance of the relationship between jobs and education and whether a different education program would have been preferable and questions on individual skills.

Section I: General Questions
These questions relate to other demographic and social questions (e.g. marital status, disability) and any training obtained through HRDC programs. As well, respondents' addresses and telephone numbers are validated and those of alternative contact persons are obtained for possible use in a subsequent survey.

4.2 Uses

This survey extends the existing base of information on the labour-market experiences of recent graduates, as in previous surveys. Information derived from the survey has the potential to shed light on many areas of

current interest. The following are examples of uses to which the survey's data may be applied.

- HRDC uses the survey data to identify potential occupational supply shortages and as basic input for job and career counselling with Canada Employment Centre (CEC) clients. These programs will also benefit from analyses of data collected on labour market experiences as they relate to development of careers and respondents' subjective assessments of jobs and relevance of their training. In such analyses, it will be possible to compare completers of trade-vocational programs with other graduates, during the five years after graduation. Moreover, job equity programs will receive important labour-market-related longitudinal information on designated groups such as women, aboriginal peoples, persons with disabilities and persons in a visible minority.

- In particular HRDC uses the survey's data to update the occupational supply and demand models and the student flow model. These models project supplies of labour by occupation and industry, especially in highly-skilled and highly-qualified categories. The models are used by HRDC in consultation with provincial governments for the development and implementation of labour market programs and policies.

- The survey provides concrete information regarding graduates' labour market experiences and career development during the five years after graduation, to aid post secondary education course selection and career counselling. The data are used at all levels of government and within the education system for occupational career counselling to help young Canadians make the appropriate career and employment choices.

5.0 Survey Methodology

The Follow-up of 1990 Graduates Survey was conducted in May and June, 1995 on the respondents to the Survey of 1990 Graduates, using a computer-assisted telephone interview (CATI) methodology. This was the first instance of using CATI techniques in the series of national graduates surveys conducted by Statistics Canada.

5.1 Population Coverage

The survey's base population is the set of graduates from Canadian postsecondary education institutions who completed the requirements for degrees, diplomas, or certificates during the calendar year 1990. More specifically, these include:

1. graduates of university programs leading to bachelors, masters or doctorate degrees, or specialized certificates/diplomas;

2. graduates of postsecondary programs (i.e. programs of one year duration or longer which normally require secondary school completion or its equivalent for admission) in the CAATs, CEGEPs, community colleges, technical schools, and similar institutions; and

3. graduates of skilled trades (i.e. pre-employment) programs which normally were 3 months or more in duration.

The population excludes:

a) graduates from private postsecondary education institutions (e.g. institutions such as commercial secretarial schools, commercial computer programming schools, etc. that do not follow a standard curriculum as established for publicly funded institutions);

b) those who completed "continuing education" courses at universities and colleges (unless these led to degrees or regular diplomas or certificates);

c) those persons who took part-time trade courses (e.g. adult education evening courses) while employed full-time;

d) persons who completed vocational programs
   i) lasting less than three months, or
   ii) other than in the skilled trades (e.g. basic training and skill development); and

e) persons in apprenticeship programs.

The list or "frame" of 1990 graduates was created, starting with a list of all universities, colleges and trade-vocational schools in Canada provided by Statistics Canada's Centre for Education Statistics. For the trade-vocational frame, a list was also obtained from HRDC of those who had taken trades training arranged for by the Department. This necessitated a check for duplicates with lists provided directly by institutions, but yielded a total list that was more complete.

Files of graduates, preferably in machine-readable form, were requested from each institution. In a few cases, files were supplied to us by provincial Ministries of higher or advanced education for all or most of the institutions in a province. For each graduate we requested his/her name, permanent address and telephone number, local address and telephone number, qualification obtained in 1990, major field of study, date of birth, student number and whether or not the graduate took his/her studies as a visa student. This procedure was followed for all provinces except Québec.

For Québec, the provincial privacy commissioner's office (Commission d'accès à l'information) ruled that we should only obtain a sample of university and college graduates, which required negotiation with the Ministère de l'Enseignement supérieur et de la Science (Ministry of Higher Education) to obtain it. Unfortunately, despite all efforts, the sample for Québec university graduates included only those who had obtained a Bachelor's, Master's or Doctorate degree. That is, it did not include any graduates with certificates or diplomas below the bachelor level ("diplômes de premier cycle"), or with certificates or diplomas above the bachelor level ("diplômes de deuxième cycle"). Since they are missing, the effective population for the survey has become the set of graduates from Canadian public postsecondary education institutions who completed the requirements for degrees, diplomas or certificates during calendar 1990. but excluding those university graduates obtaining certificates or diplomas below or above the bachelor level in Québec.

5.2 Sample Design

The Survey of 1990 Graduates and the Follow-up of 1990 Graduates were based on a stratified one-stage systematic random sample design.

5.2.1 Stratification

The population of 1990 graduates was stratified first by province. Within each province, the graduates were stratified into five levels and nine fields of study for university and career/technical programs and ten fields of study for the trade/vocational programs.

The five levels are:

1. skilled trades;

2. college;

3. undergraduate (degrees, diplomas, and certificates);

4. masters level (degrees, diplomas, and certificates);

5. doctorate.

The nine fields of study for university and career/technical programs and ten fields of study for trade/vocational programs were based on the 5-digit USIS (University Student Information System) and CCSIS (Community College Student Information System) major field of study codes. These fields of study codes with stratum codes are presented in the following tables A, B, C.

A. Major Field of Study Strata for University Programs

Stratum Code	Stratum Description	Corresponding USIS Codes
01	No specialization/ Specialization Unknown	0xxxx
02	Education	1xxxx
03	Fine Arts Humanities	2xxxx 3xxxx
04	Commerce Law Economics	412xx 427xx 433xx
05	Other Social Sciences	all other 4xxxx
06	Agricultural and Biological Sciences	5xxxx
07	Engineering	6xxxx
08	Medical and Health Professionals	7xxxx
09	Mathematical and Physical Sciences	8xxxx

B. Major Field of Study Strata for Career/Technical Programs

Stratum Code	Stratum Description	Corresponding CCSIS Codes
01	No specialization/ Specialization Unknown Arts Humanities	blank, 00000, 92000, 99999 1xxxx, 2xxxx 3xxxx
02	Health and Related Sciences	4xxxx
03	Chemical Technologies Transportation Technologies General Engineering Aeronautical Engineering Industrial Engineering	51xxx 54xxx 551xx 554xx 555xx
04	Electrical and Electronic Technologies Mathematics and Computer Science	53xxx 53xxx
05	Mechanical Engineering Architectural and Construction Engineering	552xx 553xx
06	Natural Sciences and Primary Industries	6xxxx
07	Social Sciences and Services	7xxxx
08	Secretarial Sciences Merchandising and Sales Service Industry Technologies Miscellaneous	80xxx, 81xxx 83xxx 84xxx 9xxxx, excl. 92000 and 99999
09	Management and Administration	82xxx

Note: Graduates of university transfer programs were not included in the survey.

C. Major Field of Study Strata for Trade/Vocational Programs

Stratum Code	Stratum Description	Corresponding CCSIS Codes
01	No/Unknown Specialization Arts Transportation Technology Merchandising and Sales Service Industry Technology	00000 1xxxx, 2xxxx 54xxx 83xxx 84xxx
02	Health Services and Related Social Sciences and Services	4xxxx 7xxxx, excluding 75xxx
03	Electrical and Electronic Technologies	52xxx
04	Automotive Mechanics	5523x
05	Other Mechanical	552xx, excluding 5523x
06	General Engineering Architectural and Construction Engineering	5511x 553xx, excluding 55310
07	Engineering Technologies Chemical Technology Architectural Design/Drafting Technology Industrial Engineering	50xxx, 550xx, 551xx excluding 5511x 51xxx 55310 554xx, 555xx
08	Natural Sciences and Primary Industries	6xxxx
09	Journalism Secretarial Science	31xxx 81xxx
10	Mathematics and Computer Science Business and Commerce Management and Administration	53xxx 800xx 82xxx

Note: It was felt that there were not true trade-vocational programs in the CCSIS 554xx series (Aeronautical Engineering) and the 75xxx series (Personal Development). Courses in the first series were more correctly career/technical; and the 75xxx class included Basic Training and Skill Development (BTSD), job-readiness, work-adjustment, and so on, and were intentionally omitted entirely from the survey.

5.2.2 Sample Allocation and Selection

The sample allocation to the strata was made to allow analysis at acceptable levels of detail with acceptable reliability for all provinces, levels and groups of fields of study, as defined for the 1984 and 1988 Graduate Surveys.

An independent systematic random sample of allocated size was selected from each stratum.

It should be noted that the sample for Québec university graduates provided by the Ministère de l'enseignement supérieur (Ministry of Higher Education) included only those who had obtained Bachelor's, Master's or Doctorate degrees. That is, it did not include any graduates with certificates or diplomas below the bachelor level ("diplômes de premier cycle"), or with certificates or diplomas above the bachelor level ("diplômes de deuxième cycle").

5.3 Sample Size

The tables in Section 8.1 provide the sample sizes and the number of responses by province and level of qualification for the Survey of 1990 Graduates and the Follow-up of 1990 Graduates. Detailed notes on these sample sizes and number of responses are provided at the beginning of the section.

6.0
Data Collection

6.1 Survey of 1990 Graduates

All regional office staff who worked on the survey (i.e. project supervisors, senior interviewers and interviewers) were given a training session of 1-1/2 to 2 days designed to familiarize them with the purpose and concepts of the survey, the forms and procedures involved and some basic techniques of telephone interviewing.

Data were collected using a centralised telephone interview facility in each of the Statistics Canada regional offices, between Saturday June 6 and Saturday July 4, 1992. In most offices, telephoning was conducted in two shifts, between 9:00 a.m. and 9:00 p.m.. Interviewers attempted to contact all respondents, initially using telephone numbers supplied by institutions. To aid in further tracing, a variety of other information was obtained from institutions and provided to interviewers on a separate Information Sheet. This included:

permanent address
local address
institution name and student ID number
sex
date of birth
visa-student status.

Respondents were approached in the principal official language of the institution (as far as that could be determined). Bilingual interviewers were used where required.

Using the information sheets, interviewers attempted to contact all respondents located in Canada. Tracing methods and sources included: telephone directories, city directories, alumni lists, professional associations, local taxation offices, motor vehicle licence bureaus. If an individual was found to be living in an area under the jurisdiction of a Regional Office different from the one making the initial call(s), and could not be traced by the initial Regional Office through the information supplied, the Information Sheet and a record of tracing attempts was transferred to the Regional Office in whose territory the respondent was thought to be living. Each Regional Office had a complete set of telephone directories, city directories, etc. for its territory, and usually little or nothing for cities or provinces outside its territory. It was anticipated therefore, that the receiving Regional Office might have better success in locating such hard-to-find respondents. Individuals who had left the country were excluded from the survey.

In the sample-selection process, each respondent was assigned a unique 7-digit "Respondent Number" that was printed on the Information Sheet together with the name, address, telephone number, etc. Interviewers were required to transcribe this number onto a blank questionnaire before completing on that questionnaire any further details of the interview or attempted interview. This transcription step was vital as it provided the only means to link interview responses with respondent characteristics from the sample file.

Participation in the survey was voluntary. If a respondent refused to provide some or all information requested, interviewers' supervisors were instructed to make a second call in an attempt to obtain the information. If the respondent was temporarily away or there was some language or other difficulty preventing an interview, interviewers were instructed to call back at another time. Proxy responses on behalf of the respondents were not allowed.

A single questionnaire was used to interview both university/college and trade/vocational graduates. Respondents were not expected to consult documents or files in order to respond to the questions.

Although there were a total of 178 questions in the eight sections of the questionnaire, most respondents answered only a portion of the questions within each section. Some respondents may have skipped sections because the questions were not applicable to their recent education or employment experiences.

Several provincial government Ministries of Education, and some institutions, also carried out surveys among postsecondary institution graduates. Interviewers were advised, if challenged, to explain that this survey had no connection with the other surveys and that different questions were being asked.

After all attempts to trace or interview a selected graduate were exhausted, interviewers coded the final result of the interview on the front of the questionnaire, using one of the final status codes in the following table:

CODE	NUMBER	PERCENT	STATUS
01 & 02	36,280	71.0	Contacted and completed or partial interview
03	593	1.2	Refusal
04	377	0.7	Already contacted (duplicate)
05	1,344	2.6	Absent for duration of survey
06	381	0.8	Unlisted telephone number
07	627	1.2	No answer (after several tries)
08	524	1.0	Can't be reached by telephone
09	4,356	8.5	Unable to trace
10	1,880	3.7	Interview ended at item 2
11	1,823	3.6	No longer living in Canada
12	56	0.1	Deceased
13	1,920	3.8	Other (this includes respondents given a response code of 01 or 02, but with no questionnaire)
14	950	1.9	Said "no" to H21
TOTAL	**51,111**	**100.0**	

It should be noted that these counts are from final status codes assigned by interviewers. The counts were revised during processing.

The total sample for the Survey of 1990 Graduates and the corresponding number of respondents were distributed according to province/territory of institutions as follows:

Province/Territory	Total Sample	Total Respondents
Newfoundland	2,468	1,902
Prince Edward Island	1,165	948
Nova Scotia	3,271	2,613
New Brunswick	2,691	2,269
Quebec	9,092	6,591
Ontario	14,340	9,837
Manitoba	3,264	2,602
Saskatchewan	3,030	2,338
Alberta	5,024	3,292
British Columbia	5,954	3,639
Yukon	199	93
North West Territories	613	156
Canada	**51,111**	**36,280**

6.2 Follow up of 1990 Graduates

Data were collected by CATI through the Statistics Canada regional offices from April 24 to July 15,1995. Interviewers attempted to contact all respondents to the 1992 survey, using the addresses and telephone numbers obtained in the June 1992 interview as well as a selection of 1992 results necessary for the proper flow of the questionnaire.

Any respondent no longer living at the address and telephone number obtained in June 1992 was traced again through enquiry at last known telephone numbers, permanent residence telephone numbers (e.g., parents' home), alumni lists, and so on.

Respondents were approached in the official language identified in the 1992 interview. Bilingual interviewers were used where required.

Interviewers were instructed to ask the questions directly to the respondent. No proxy interviews were allowed.

Although there were nine sections of questions, most respondents only answered a portion of the questions within each section. Some respondents may have skipped sections because the questions were not applicable to their recent education or employment experiences.

CATI methodology also eliminated the need for paper questionnaires. The paper document was produced only as a reference document and for the training of interviewers.

Several provincial government Ministries of Education, and some institutions, also carried out surveys among postsecondary institution graduates. Interviews were advised, if challenged, to explain that the Follow-up survey had no connection with the other surveys and that different questions were being asked.

The total sample size for the Follow-up of 1990 Graduates and the corresponding number of respondents were distributed according to province/territory of institutions as follows:

Province/Territory	Total Sample	Total Respondents
Newfoundland	1,902	1,716
Prince Edward Island	948	837
Nova Scotia	2,613	2,316
New Brunswick	2,269	1,954
Quebec	6,591	5,522
Ontario	9,837	7,899
Manitoba	2,602	2,279
Saskatchewan	2,338	2,121
Alberta	3,292	2,888
British Columbia	3,639	3,081
Yukon	93	81
North West Territories	156	122
Canada	**36,280**	**30,816**

7.0 Data Processing

7.1 Survey of 1990 Graduates

7.1.1 Data Capture

All forms were data captured in July and August 1992 using mini-computers in each of Statistics Canada's Regional Offices. This took approximately 4 weeks. The data capture program allowed for a valid range of codes for each question and automatically followed the flow of the questionnaire. The "Respondent number" uniquely identifying each respondent was automatically checked against a look-up table to guard against capture errors. After data capture, an unedited version of all captured information was electronically transmitted to Ottawa for the creation of an initial computer file.

7.1.2 Edit and Imputation

The first stage of survey processing involved the linkage of 5 captured screens for each document into single records. The file was verified for duplicates and missing screens. Duplicates were dropped.

The next stage involved the reformation of data by validating cell values and transforming these keyed values into new label values. The data underwent further reformation by the destringing of certain data (e.g. all multiple response questions) into fixed positional arrays and the standardization of values into common formats (e.g. all date values to year/month/day). Records were then split into a response file and a non-response file. Respondents who did not wish to share their answers with other government departments were included in the non-response file.

Text fields were removed from the response records and replaced by flags. Based on text categories (e.g. industry, occupation, education, or other “specify” entries), files were constructed for further processing during the coding stages.

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

The final stage of the editing process involved editing all response survey records according to pre-specified edit rules to check for errors, gaps, and inconsistencies in the survey data. Checks were made to insure that numerical answers to certain questions fell within acceptable logical ranges. Checks were also made to ensure proper flows through the skip patterns of the questionnaire. When errors or inconsistencies were found, the erroneous information was either blanked out or replaced by "not stated" values. Editing was mostly “top-down” meaning that when a flow question was encountered, the flow pattern indicated by the response of that question was accepted as true. Records which were judged to have insufficient or irreconcilable data were removed from the file.

Imputation was not appropriate for most items and thus "not stated" codes were usually assigned for missing data. For a very few records, the respondent's sex was not provided by the institution and could not be determined from the given name. It was therefore coded "not stated".

7.1.3 Coding of Open-ended Questions

A number of data items on each questionnaire were recorded in an open-ended format. Some of these were subsequently coded. These were (I) industry and occupation of the respondent at certain reference periods, (ii) major field *of* study descriptions, and (iii) other "specify" categories.

7.1.4 Coding of Industry and Occupation

For each job held by the respondent in the reference periods, the questionnaire collected information on the name of the employer, the kind of business, industry or service the employer was in, the kind of work done and the usual duties or responsibilities of the respondent in the job. This information was used to assign industry and occupation codes to each job using the 1980 version of Statistics Canada's Standard Industrial Classification (SIC) and Standard Occupational Classification (SOC). SIC codes were assigned to the 3-digit level, and SOC codes to the 4-digit level (see Appendix A).

The information on the name of the employer, etc. was first put through the automated coding system developed for use by the Labour Force Survey. Where a match to the database of descriptions was obtained, the appropriate industry and/or occupation code was assigned. The remaining partially coded or uncoded records were coded manually by a staff of experienced coders. All coded values were then linked to the survey data file.

7.1.5 Coding of Field of Study Descriptions

Field of study descriptions were coded under two different coding systems, the University Student Information System (USIS) and the Community College Student Information System (CCSIS) "Spemaj" 5 digit field of study codes. The choice of coding system for each description was determined from corresponding questions on program levels within institutions. The descriptions were then passed through an automated coding system developed internally for this survey. Descriptions that matched the database were automatically assigned codes whereas the unmatched descriptions were coded by staff involved in data processing operations. After a final quality review the codes were matched back to the data file.

7.1.6 Coding of “Other, specify” Categories

The responses to questions containing an “Other, specify” category were verified separately to determine the validity of each response. Entries were manually reviewed to determine whether they belonged in the “Other” category or whether they should have been included in one of the listed response categories. Any response which fit into an existing category was recoded to the appropriate category. In some cases, only part of the "Other, specify" could be recoded. When all the corrections had been completed, the files containing the corrected responses were matched back to the data file.

7.1.7 Creation of Derived Variables

A number of data items on the microdata file were derived by combining similar items from the questionnaire to create such variables as labour force status, relationship of jobs to education, class of worker, age of respondent, number of children, etc. More complete descriptions are provided for each derived variable in the Record Layout.

7.2 Follow-up 1990 Graduates

7.2.1 Data Capture

The F90G survey data were collected using a CATI application known as CASES (Computer Assisted Survey Execution System). The CASES system has two main parts; Case Management and a part that is specific to each survey.

The Case Management system controls the case assignment and data transmission for the survey. For the F90G, a case referred to an individual graduate selected for the F90G sample. The Case Management system also automatically recorded 'survey management' information for each contact (or attempted contact) with respondents, and provided reports on the progress of the collection process throughout the entire interview period.

The survey-specific part of CASES included an introductory component with procedures for contact and verification of graduates that were selected for interview. Once contact had been made, the CASES system generated the questionnaire components for the interview with the respondent. The interviewer asked the respondent the questions, then entered the responses into the computer as the interview progressed. The CATI program also performed on-line edits allowing for correction of potential data errors while the respondent was available on the telephone.

As cases were completed, most of which were completed interviews, data files containing the records of all captured information were electronically transmitted to Ottawa. This usually occurred on a daily basis, resulting in the creation of a raw data file to be used in post-survey processing.

7.2.2 Edit and Imputation

The raw survey data coming out of the F90G CASES application were generally cleaner than the previous S90G data collected using a paper-and-pencil telephone interview collection method. Nevertheless, some post-collection editing was still necessary to ensure that users of the final microdata file could clearly distinguish between valid responses, item non-response and valid skips in survey questions. This was necessary in part due to the need to transform certain data values to 'numeric' character values from 'alpha' characters, as they had initially been entered in the CASES system.

At an early stage in processing, survey records were split into a response file and a non-response file. Respondents who did not wish to share their answers with other government departments were included in the non-response file.

Preliminary grooming of the response file included changing the response values for categorical questions from '3's to '2's (a 'no' response), 'x's to '7's (a 'don't know' response) , 'r's to '8's (a ' refusal'), and blanks to valid skips (6, 96, 996, etc.). Various field lengths were increased to correspond to expanded response categories for some questions. Text was removed where open-ended questions were asked and replaced by 1 character flags. Industry, occupation, education and 'other, specify' entry files were created for further processing during the coding stages.

The final stage of the edit process involved editing the response file records according to pre-specified edit rules to check for errors, gaps, and inconsistencies in the survey data. Checks were made to insure that numerical answers to certain questions fell within acceptable logical ranges. Checks were also made to ensure proper flows through the skip patterns of the questionnaire.

Editing was mostly 'top-down' meaning that when a flow question was encountered, the flow pattern indicated by the response to that question was accepted as true. The question answer paths were cleaned for every 'go to' encountered from a legitimate answer and all the subsequent questions were set to valid skip. The next processing step handled 'don't know's' or 'refusal's', whereby the subsequent questions were set to 'not stated' in that path flow.

As was the case in the S90G processing, imputation was not appropriate for most items and thus 'not stated' codes were usually assigned for missing data. Records which were judged to have insufficient or irreconcilable data were removed from the file.

In cases where some data fields in the F90G were identified as having the same variable name as in the S90G, they were renamed to a unique field name. These include such variables as B17 to B17F, B18 to B18F, the 'F' signifing a 'Follow-up' variable. Refer to the Record Layout for a detailed list.

To ensure consistency in the longitudinal file containing both S90G and F90G data, and due to the edit process for question skip pattern values used for the F90G file, all blank values on the S90G file portion were converted to codes of 6, 96, 996, etc. for valid skips. Finally, the processing of some derived variables for the F90G required the use of a number of variables from the S90G file.

7.2.3 Coding of Open-ended Questions

As in the S90G survey, a number of data items on the F90G CATI questionnaire were recorded in an open-ended format. Some of these were subsequently coded. These were (I) industry and occupation of the

respondent at certain reference periods, (ii) major field of study descriptions, and (iii) other "specify" categories.

7.2.4 Coding of Industry and Occupation

For each job held by the respondent in the reference periods, the questionnaire collected information on the name of the employer, the kind of business, industry or service the employer was in, the kind of work done and the usual duties or responsibilities of the respondent in the job. This information was used to assign industry codes to each job using the 1980 version of Statistics Canada's Standard Industrial Classification (SIC).

SIC codes were assigned to the 3-digit level. The range of values was from 000 to 999. To allow for unique and consistent coding, the SIC was expanded to 4 digits to accommodate for a valid skip code of 9996. The range of values is now 0000 to 9999.

Occupation codes were assigned using both the 1980 and 1991 versions of Statistics Canada's Standard Occupational Classification (SOC). 1980 SOC codes were assigned to the 4-digit level, while the 1991 SOC also includes the National Occupational Classification (NOC). 1991 SOC\NOC codes were assigned to the 4-digit level each with a decimal in the fifth position for a total length of nine (ANNN.NNNN). The 1991 SOC is in the first 4 positions and the 1991 NOC is in the last 4 positions, separated by the decimal.

The information on the name of the employer, etc. was first put through the automated coding system developed for use by the Labour Force Survey. Where a match to the database of descriptions was obtained, the appropriate 1980 SIC code and/or 1980 SOC code was assigned. The remaining partially coded or uncoded records were coded manually by a staff of experienced coders. All 1991 SOC codes were assigned manually by these same coders since the automated coding system was not available for the 1991 codes.

All coded records were then linked to the survey data file. In cases where the occupation descriptions were incomplete but could be coded to a major group level, partial codes were completed. A partial code contains an “X” in the minor and/or unit group level where the complete code could not be assigned. This occurred more frequently with the 1991SOC\NOC codes.

Finally, for the S90G portion of the longitudinal file, the 1991 SOC\NOC codes were added to all the previously coded 1980 SOC records. Thus the file is now completely coded to the 1980 SIC\SOC and the 1991 SOC\NOC standards.

Detailed lists of all the SIC\SOC\NOC codes can be found in Appendix A.

7.2.5 Coding of Field of Study Descriptions

Similar to the S90G, field of study descriptions were coded under two different coding systems, the University Student Information System (USIS) and the Community College Student Information System (CCSIS) 'Spemaj' 5 digit field of study codes. The choice of coding system for each description was determined from corresponding questions on program levels within institutions. The descriptions were then passed through an automated coding system developed internally for this survey. Descriptions that matched the database were automatically assigned codes whereas the unmatched descriptions were coded by staff involved in data processing operations. After a final quality review the codes were matched back to the data file.

A second set of field of study codes, called harmonization codes, was also completed on all USIS\CCSIS field of study codes for both the S90G and F90G survey files. The harmonization codes standardize the USIS or CCSIS codes into minor groups (102 categories) or major groups (11 categories) for more universal use in field of study comparisons with other datasets.

See Appendix B for more details on all the field of study code sets.

In the F90G, questions such as HA4 and HG2 were structured differently in terms of the response categories compared to the level of education questions in B20 and C8. For example, it was particularly difficult to assess the level of education for categories like professional association diploma, certificate or license such as in accounting, banking or insurance. In some cases, in order to determine a 'USIS' or 'CCSIS' codebook designation and improve the quality and accuracy of the field of study codes, respondents were linked back to the S90G for their reported certification level. Furthermore, this improved the data quality of the field of study harmonization coding completed on the longitudinal file.

7.2.6 Coding of 'Other, specify' Categories

This process was similar to the strategy used in the S90G file.
The responses to questions containing an 'Other, specify' category were verified separately to determine the validity of each response. Entries were manually reviewed to determine whether they belonged in the 'Other, specify' category or whether they should have been included in one of the listed response categories. Any response which fit into an existing category was recoded to the appropriate category. In some cases, only part of the 'Other, specify' could be recoded.

A new code was created if a particular response occurred enough times to match the lowest frequency in the given list of responses or if it represented approximately 10% of the total text write-ins. For example, in the F90G, questions A6, B30, B33, C21 and C24 asked for the reason a respondent worked less than 30 hours per week. Four new codes were created for these variables after analysis of the responses.
These were:

08 - Business conditions; nature of the job
(e.g. downsizing, seasonal, on call)
09 - Contract
10 - Shortage of work; no funding (e.g. layoffs, budget cuts)
11 - Have 2 or more jobs

When all the corrections had been completed, the files containing the corrected responses were matched back to the main data file.

7.2.7 Creation of Derived Variables

A number of data items on the microdata file were derived by combining similar items from the F90G questionnaire and the S90G file to create such variables as labour force status, relationship of jobs to education, class of worker, age of respondent, number of children, etc. More complete descriptions are provided for each derived variable in the Record Layout.

7.3 Weighting

The principle behind estimation in a probability sample such as the Survey/Follow-up of 1990 Graduates 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 received a masters degree in 1990 in Canada is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

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

During processing of the 1992 survey results it was discovered that several institutions had sent us files of their graduates that were substantially of 1991 graduates (e.g., they had included 1990/91 academic year graduates rather than 1990 calendar year). This error had not been detected during processing of their files to compile the survey frame, prior to sampling. It was found only at the time of interview when respondents answered "no" to question A3 of the questionnaire (see section 12.1) (i.e., they had not graduated in 1990). These respondents were given an interview final-status code 10 by interviewers, i.e. "interview ended at item A3 ..." (see section 6.1). This code meant that the respondents would have been classed as outside the population of interest, and not included in the non-response correction of the weights. However, given the magnitude of the effect on regional estimates, we chose to reclassify the final response for these respondents as "other" (code 13). There are two implied assumptions in this procedure:

1. that the counts of true calendar-1990 graduates would not be substantially different from the counts in the mixed 1990-1991 files we had received; and

2. that the responses we obtained from true 1990 graduates were representative of all 1990 graduates potentially in the sample, i.e., including those we should have obtained.

7.4 Suppression of Confidential Information

In order to avoid duplication of enquiry, Statistics Canada conducted both the Survey of 1990 Graduates and the Follow-up of 1990 Graduates with Human Resources Development Canada, the Department of the Secretary of State, and the provincial Ministries of Education and Labour. However, the information on the microdata files provided to these departments does not contain respondents' names or other identifying data (e.g. names of employers) and is kept confidential and used only for statistical purposes.

If respondents indicated that they did not wish to share their answers to the survey questions as per the above 'data-sharing' agreement, or their response to the 'data-sharing' question was blank, the respondents' information was removed from the file and they were considered to be non-response.

8.0

Data Quality

8.1 Response Rates

The following tables present a summary of the number of graduates in the subpopulations and selected in the sample for each of the 5 subpopulations in each province/territory, and the response rates achieved for both the Survey of 1990 Graduates and for the Follow-up of 1990 Graduates.

Subpopulation size: refers to the number of 1990 graduates reported by the institutions included in the survey. Some institutions reported too late or did not report at all, and some reporting institutions did not report all graduates.

Selected sample size: refers to the number of 1990 graduates actually selected for the survey in 1992.

"Responding" sample size: refers to the number of graduates who at the time of the 1992 survey: (a) were still living in Canada and provided complete or partial information (i.e., with final status codes 01 or 02); (b) who turned out not to have graduated at all in 1990, or to have taken a trade-vocational course of less than three months, thus were in the frame in error (final status code 10); (c) were no longer living in Canada or were dead (final status codes 11 or 12); or (d) indicated upon subsequent contact that they graduated with more than one degree, diploma or certificate in 1990 (final status code 04).

Domain of interest: refers to those 1990 graduates from the responding sample who fall into category (a) above, i.e. who were still living in Canada in June-July 1992, and provided complete or partial information. Column 6 in the table shows the percent of the selected sample in the domain of interest. The relatively low percentage for doctorate graduates is a reflection of the high proportion that were no longer living in Canada in 1992.

Actual Sample Size '95: refers to the number of 1990 graduates from the "domain of interest" described above, who agreed to share their data in June-July 1992 and were still living in Canada in May-June 1995.

Actual Responses to '95: refers to the number of 1990 graduates from the actual sample for 1995 who provided complete or partial information and agreed to share it in May-June 1995. The final column in the table shows the percentage of the actual responses in the actual sample for 1995.

Note:
It should be noted that the sample for Québec university graduates provided by the Ministère de l'enseignement supérieur (Ministry of Higher Education) included only those with Bachelor's, Master's and Doctorate degrees. That is, it did not include any graduates with certificates or diplomas below the bachelor level ("diplômes de premier cycle"), or with certificates or diplomas above the bachelor level ("diplômes de deuxième cycle").

FOLLOW-UP OF 1990 GRADUATES

REPORTED SUBPOPULATION SIZES, SELECTED AND RESPONDING SAMPLE SIZES, RESPONSE RAT
DOMAIN OF INTEREST BY PROVINCE/TERRITORY OF INSTITUTION AND LEVEL OF CER

Province/Territory Level of Certification	(1) Reported Subpopu-lation Size	(2) Selected Sample Size	(3) Responding Sample Size 1992	(4) Response Rate 1992 (3)/(2) X 100%	(5) Domain of Interest 1992	(6) Domain Percent (5)/(2) X 100%	(Ac Sa Size
NEWFOUNDLAND							
Bachelor's/1st Prof. Deg.	2324	801	716	89.4%	679	84.8%	
Master's	196	196	165	84.2%	157	80.1%	
Doctorate	18	18	15	83.3%	13	72.2%	
Career/Technical	807	590	493	83.6%	479	81.2%	
Trade/Vocational	1475	863	660	76.5%	574	66.5%	
Total	4820	2468	2049	83.0%	1902	77.1%	
PRINCE EDWARD ISLAND							
Bachelor's/1st Prof. Deg.	455	347	330	95.1%	315	90.8%	
Master's	10	10	9	90.0%	9	90.0%	
Doctorate	-	-	-	-	-	-	
Career/Technical	602	412	376	91.3%	338	82.0%	
Trade/Vocational	396	396	351	88.6%	286	72.2%	
Total	1463	1165	1066	91.5%	948	81.4%	
NOVA SCOTIA							
Bachelor's/1st Prof. Deg.	5603	1056	944	89.4%	886	83.9%	
Master's	794	577	520	90.1%	465	80.6%	
Doctorate	63	63	59	93.7%	46	73.0%	
Career/Technical	786	489	461	94.3%	417	85.3%	
Trade/Vocational	4167	1086	921	84.8%	799	73.6%	
Total	11413	3271	2905	88.8%	2613	79.9%	
NEW BRUNSWICK							
Bachelor's/1st Prof. Deg.	2932	885	827	93.4%	803	90.7%	
Master's	370	370	335	90.5%	296	80.0%	
Doctorate	26	26	26	100.0%	19	73.1%	
Career/Technical	746	494	460	93.1%	450	91.1%	
Trade/Vocational	2384	916	820	89.5%	701	76.5%	
Total	6458	2691	2468	91.7%	2269	84.3%	

REPORTED SUBPOPULATION SIZES, SELECTED AND RESPONDING SAMPLE SIZES, RESPONSE RATES, AN
INTEREST BY PROVINCE/TERRITORY OF INSTITUTION AND LEVEL OF CERTIFICA

Province/Territory Level of Certification	(1) Reported Subpopu-lation Size	(2) Selected Sample Size	(3) Responding Sample Size 1992	(4) Response Rate 1992 (3)/(2) X 100%	(5) Domain of Interest 1992	(6) Domain Percent (5)/(2) X 100%	(Ac Sa Size
QUEBEC							
Bachelor's/1st Prof. Deg.	25528	2913	2394	82.2%	2324	79.8%	
Master's	5015	2021	1643	81.3%	1494	73.9%	
Doctorate	689	689	565	82.0%	436	63.3%	
Career/Technical	15909	1965	1566	79.7%	1539	78.3%	
Trade/Vocational	5853	1504	1062	70.6%	798	53.1%	
Total	52994	9092	7230	79.5%	6591	72.5%	
ONTARIO							
Bachelor's/1st Prof. Deg.	49001	3397	2780	81.8%	2599	76.5%	
Master's	7384	2762	2245	81.3%	2005	72.6%	
Doctorate	1159	1159	948	81.8%	748	64.5%	
Career/Technical	23968	3591	2845	79.2%	2699	75.2%	
Trade/Vocational	17118	3431	2234	65.1%	1786	52.1%	
Total	98630	14340	11052	77.1%	9837	68.6%	
MANITOBA							
Bachelor's/1st Prof. Deg.	5038	1053	919	87.3%	859	81.6%	
Master's	622	622	549	88.3%	490	78.8%	
Doctorate	72	72	69	95.8%	47	65.3%	
Career/Technical	1416	681	605	88.8%	588	86.3%	
Trade/Vocational	2088	836	696	83.3%	618	73.9%	
Total	9236	3264	2838	86.9%	2602	79.7%	
SASKATCHEWAN							
Bachelor's/1st Prof. Deg.	4789	969	849	87.6%	809	83.5%	
Master's	506	506	442	87.4%	389	76.9%	
Doctorate	70	70	60	85.7%	40	57.1%	
Career/Technical	1333	690	570	82.6%	531	77.0%	
Trade/Vocational	1763	795	645	81.1%	569	71.6%	
Total	8461	3030	2566	84.7%	2338	77.2%	

FOLLOW-UP OF 1990 GRADUATES

REPORTED SUBPOPULATION SIZES, SELECTED AND RESPONDING SAMPLE SIZES, RESPONSE RATES, AN
INTEREST BY PROVINCE/TERRITORY OF INSTITUTION AND LEVEL OF CERTIFICA

Province/Territory Level of Certification	(1) Reported Subpopu-lation Size	(2) Selected Sample Size	(3) Responding Sample Size 1992	(4) Response Rate 1992 (3)/(2) X 100%	(5) Domain of Interest 1992	(6) Domain Percent (5)/(2) X 100%	(Ac Sa Size
ALBERTA							
Bachelor's/1st Prof. Deg.	8872	1154	1015	88.0%	963	83.4%	
Master's	1174	708	609	86.0%	551	77.8%	
Doctorate	280	280	226	80.7%	155	55.4%	
Career/Technical	7262	1421	919	64.7%	892	62.8%	
Trade/Vocational	4992	1461	897	61.4%	731	50.0%	
Total	22580	5024	3666	73.0%	3292	65.5%	
BRITISH COLUMBIA							
Bachelor's/1st Prof. Deg.	9104	1186	937	79.0%	864	72.9%	
Master's	1627	938	772	82.3%	656	69.9%	
Doctorate	269	269	220	81.8%	143	53.2%	
Career/Technical	6618	1302	924	71.0%	872	67.0%	
Trade/Vocational	12070	2259	1224	54.2%	1104	48.9%	
Total	29688	5954	4077	68.5%	3639	61.1%	
YUKON							
Career/Technical	30	30	23	76.7%	20	66.7%	
Trade/Vocational	169	169	102	60.4%	73	43.2%	
Total	199	199	125	62.8%	93	46.7%	
NORTHWEST TERR.							
Career/Technical	98	98	73	74.5%	58	59.2%	
Trade/Vocational	515	515	302	58.6%	98	19.0%	
Total	613	613	375	61.2%	156	25.4%	
CANADA							
Bachelor's/1st Prof. Deg.	113646	13761	11711	85.1%	11101	80.7%	
Master's	17698	8710	7289	83.7%	6512	74.8%	
Doctorate	2646	2646	2188	82.7%	1647	62.2%	
Career/Technical	59575	11763	9315	79.2%	8883	75.5%	
Trade/Vocational	52990	14231	9914	69.7%	8137	57.2%	
Total	246555	51111	40417	79.1%	36280	71.0%	

8.2 Survey Errors

The estimates derived from this survey are based on a sample of graduates from Canadian postsecondary education institutions who completed the requirements for degrees, diplomas, or certificates during the calendar year 1990. In the case of the 'Doctorate' level graduates, a complete census was taken for the survey.

However, somewhat different figures might have been obtained if a complete census of all 1990 graduates 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 Data Processing

In future, when designing the questionnaire, if there is intention to code 'other, specify' texts back to existing categories it is best not to have any 'go to' flows. During processing of the F90G, the flow from questions A10F to A12 could not be reconciled because the flow was not respected. Yet all the answers in A12 were kept because they were important for determining labour force status.

8.2.2 Non-response

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, or could not recall the requested 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 (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 8.3% of 1990 graduates did not get employment in their field of study, with a standard error of 0.3%. Then the co-efficient of variation of the estimate is calculated as:

$$\left(\frac{.003}{.083}\right) \; x \; 100\% \;\; ' \;\; 3.6\%$$

9.0 Guidelines for Tabulation, Analysis and Release

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

9.1 Rounding Guidelines

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

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

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

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

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Follow-up of 1990 Graduates 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 files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

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

9.2.1 Definitions of types of estimates: Categorical vs. Quantitative

Before discussing how the Follow-up of 1990 Graduates data can be tabulated and analysed, 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 Follow-up of 1990 Graduates survey.

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 graduates that worked at a job or business in the week before they were surveyed, or the proportion of those same graduates who were paid workers during that week, 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: Last week, did you work at a job or business?
R: Yes / No / Don't know / Refused

Q: Last week, were you a paid worker or self-employed?
R: Paid worker / Self-employed / Other / Don't Know / Refused

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 $\hat{X}/\hat{\imath}$ 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 hours worked per week in their main job for graduates working at a job or business in the week before they were surveyed. The numerator is an estimate of the total number of hours worked per week in their main job for all graduates working at a job or business that week, and its denominator is the number of graduates who worked at a job or business that week.

Examples of Quantitative Questions :

Q: How many hours a week did you usually work at that job?
R: |_|_|_| hours

Q: How much do you owe to the student loan program now?
R: |_|_|_|_|_| dollars

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.

Details of how the weights are used in the calculation of the categorical estimates are presented in Chapter 11.

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 worked by graduates in their main jobs in the week before they were surveyed, multiply the value reported in B29F or C20F (hours per week worked in current main job) by the final weight for the record, then sum this value over all records with A1F=1 (worked at a job or business last week).

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 worked by graduates in their main job in the week before they were surveyed,

(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 A1F=1, then
(c) divide estimate (a) by estimate (b).

Details of how the weights are used in the calculation of the quantitative estimates are presented in Chapter 11.

9.3 Guidelines for Statistical Analysis

The Follow-up of 1990 Graduates is based upon a complex sample design, with stratification 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 poor. 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 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 Follow-up of 1990 Graduates, 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 section 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 section 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. (*The figure 30 is for use with surveys with generally small sampling fractions. From time to time, a lower figure may be appropriate for surveys with higher sampling fraction.*)

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% - 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% - 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 stratified 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.

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.

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

FOLLOW-UP OF 1990 GRADUATES
INPUT DATA FOR APPROXIMATE SAMPLING VARIABILITY TABLES

PROVINCES	UNIVERSITY			COLLEGE			TRADE/VOCATIONAL		
	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION
CANADA	2.3	16,793	133,990	1.6	7,755	59,575	1.6	6,932	52,990
NEWFOUNDLAND	1.1	755	2,538	1.0	437	807	1.1	536	1,475
P.E.I.	1.0	292	465	1.0	301	602	1.0	249	396
NOVA SCOTIA	1.3	1,245	6,460	1.0	382	786	1.1	725	4,167
NEW BRUNSWICK	1.2	984	3,328	1.0	398	746	1.0	597	2,384
QUEBEC	1.8	3,618	31,232	1.2	1,341	15,909	1.1	683	5,853
ONTARIO	2.3	4,517	57,544	1.3	2,248	23,968	1.2	1,409	17,118
MANITOBA	1.4	1,259	5,732	1.0	525	1,416	1.0	540	2,088
SASKATCHEWAN	1.4	1,147	5,365	1.0	493	1,333	1.0	522	1,763
ALBERTA	1.4	1,511	10,326	1.1	802	7,262	1.0	631	4,992
BRITISH COLUMBIA	1.6	1,465	11,000	1.1	765	6,618	1.3	899	12,070
N.W.T.	-	-	-	1.1	48	98	1.1	75	515
YUKON	-	-	-	-	-	-	1.1	66	169

PROVINCES	BACHELORS			MASTERS			DOCTORATE		
	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION	DESIGN EFFECT	SAMPLE SIZE	POPU-LATION
CANADA	1.6	9,678	113,646	1.5	5,699	17,698	1.0	1,416	2,646
NEWFOUNDLAND	1.0	599	2,324	1.2	145	196	-	11	18
P.E.I.	1.0	283	455	-	9	10	-	-	-
NOVA SCOTIA	1.0	793	5,603	1.0	415	794	-	37	63
NEW BRUNSWICK	1.0	707	2,932	1.0	260	370	-	17	26
QUEBEC	1.2	1,962	25,528	1.2	1,272	5,015	1.0	384	689
ONTARIO	1.3	2,184	49,001	1.3	1,701	7,384	1.0	632	1,159
MANITOBA	1.0	775	5,038	1.0	445	622	-	39	72
SASKATCHEWAN	1.0	748	4,789	1.0	363	506	-	36	70
ALBERTA	1.0	870	8,872	1.0	508	1,174	1.0	133	280
BRITISH COLUMBIA	1.0	757	9,104	1.1	581	1,627	1.0	127	269

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 “university graduates who were employed in temporary positions ” in the week before they were surveyed is more reliable than the estimated number of “university graduates who were employed in temporary positions” in the week before they were surveyed. (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. 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:

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

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 $F_{\hat{d}}/\hat{d}$. 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 university graduates employed” in the week before they were surveyed and the numerator is the number of “university graduates who were employed in temporary positions” in the week before they were surveyed.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of “university graduates who were employed in temporary positions” as compared to number of “university graduates who were employed in permanent positions”, 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:

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

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 $F_{\hat{R}}/\hat{R}$. 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.

10.1.1 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.

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

Suppose that a user estimates that 17,219 university graduates were employed in temporary positions in the week before they were surveyed. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the cv table for CANADA.

(2) The estimated aggregate (17,219) 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 17,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.9%.

(4) So the approximate coefficient of variation of the estimate is 2.9%.The finding that there were 17,219 university graduates who were employed in temporary positions in the week before they were surveyed is publishable with no qualifications.

Example 2 : Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates that 4172/17,219=24.2% of university graduates who were employed in temporary positions in the week before they were surveyed also reported that they worked part-time (i.e. less than 30 hours per week). How does the user determine the coefficient of variation of this estimate?

(1) Refer to the table for UNIVERSITY: CANADA.

(2) Because the estimate is a percentage which is based on a subset of the total population (i.e., university graduates employed in temporary positions who worked part-time), it is necessary to use both the percentage (24.2%) and the numerator portion of the percentage (4172) in determining the coefficient of variation.

(3) The numerator, 4172, 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 4000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 25.0%.

(4) The figure at the intersection of the row and column used, namely 5.6% is the coefficient of variation to be used.

(5) So the approximate coefficient of variation of the estimate is 5.6%. The finding that 24.2% of university graduates employed in temporary positions in the week before they were surveyed were working part-time can be published with no qualifications.

Example 3 : Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 3167/10975=28.9% of female university graduates employed in temporary positions were working part-time, while 1005/6239=16.1% of male university graduates employed in temporary positions were working part-time. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the UNIVERSITY: CANADA cv table for in the same manner as described in example 2 gives the cv of the estimate for females as 6.2%, and the cv of the estimate for males as 11.9%.

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

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

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 $\hat{d} = (.289-.161) = .128$ is:

$$F_{\hat{d}} = \sqrt{[(.289)(.062)]^2 + [(.161)(.119)]^2}$$

$$= \sqrt{(.000321) + (.000367)}$$

$$= .026$$

3) The coefficient of variation of $\hat{d}$ is given by $F_{\hat{d}}/\hat{d} = .026/.128 = 0.20$.

(4) So the approximate coefficient of variation of the difference between the estimates is 20%. This estimate can be released, but with the estimate flagged as being of "marginal" quality and with an accompanying explanation, as defined in section 9.4.

Example 4 : Estimates of Ratios

Suppose that the user estimates that 10,975 female university graduates were employed in temporary positions in the week before they were surveyed, while 6239 male university graduates were employed in temporary positions. The user is interested in comparing the estimate of women versus that of men in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

(1) First of all, this estimate is a ratio estimate, where the numerator of the estimate (= $\hat{X}_1$) is the number of female university graduates who were employed in temporary positions. The denominator of the estimate (= $\hat{X}_2$) is the number of male university graduates who were employed in temporary positions.

(2) Refer to the table for UNIVERSITY: CANADA.

(3) The numerator of this ratio estimate is 10,975. The figure closest to it is 11,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 3.7%.

(4) The denominator of this ratio estimate is 6239. The figure closest to it is 6000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 5.1%.

(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 $\bar{X}_1$ and $\bar{X}_2$ respectively.

That is ,

$$\alpha_{\hat{R}} = \sqrt{(.037)^2 + (.051)^2}$$

$$= 0.063$$

The obtained ratio of women versus men university graduates who were employed in temporary positions the week before they were surveyed is 10,975/6239, which is 1.76:1. The coefficient of variation of this estimate is 6.3%, 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:

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

where $\%_{\hat{X}}$ 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 university graduates employed in temporary positions who worked part-time(from Example 2, section 10.1) would be calculated as follows.

$\hat{X}$ = 24.2% (or expressed as a proportion = .242)

t = 2

$\%_{\hat{X}}$ = 5.6% (.056 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

CI_X = {.242 - (2) (.242) (.056), .242 + (2) (.242) (.056)}

CI_X = {.242 - .027, .242 + .027}

CI_X = {.215, .269}

With 95% confidence it can be said that between 21.5% and 26.9% of university graduates employed in temporary positions the week before they were surveyed were working part-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 2 characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $F_{\hat{d}}$.

If $t = \frac{\hat{X}_1 - \hat{X}_2}{F_{\hat{d}}}$ 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 female university graduates employed in temporary positions who worked part-time and the proportion of male university graduates employed in temporary positions who worked part-time. From example 3, section 10.1, the standard error of the difference between these two estimates was found to be = .026. Hence ,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{F_{\hat{d}}} = \frac{.289 - .161}{.026} = \frac{.128}{.026} = 4.9.$$

Since t = 4.9 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the Follow-up of 1990 Graduates 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-off's for the Follow-up of 1990 Graduates

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.

FOLLOW-UP OF 1990 GRADUATES RELEASE CUT-OFF VALUES

LEVELS	UNIVERSITY			COLLEGE			TRADE/VOCATIONAL		
	C.V.			C.V.			C.V.		
PROVINCES	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX
CANADA	693	303	171	453	198	112	448	196	111
NEWFOUNDLAND	130	58	33	63	29	16	104	47	27
P.E.I.	53	24	14	66	30	18	51	24	14
NOVA SCOTIA	244	109	62	70	32	18	221	99	57
NEW BRUNSWICK	146	65	37	64	29	17	138	62	36
QUEBEC	577	254	144	510	226	128	328	148	84
ONTARIO	1,112	490	277	505	223	126	525	233	132
MANITOBA	231	103	58	94	43	24	134	61	35
SASKATCHEWAN	237	106	60	93	42	24	117	53	30
ALBERTA	350	155	88	352	158	90	275	124	71
BRITISH COLUMBIA	437	195	110	333	149	85	610	274	156
N.W.T.	-	-	-	45	27	17	180	98	60
YUKON	-	-	-	-	-	-	64	36	22

LEVELS	BACHELORS			MASTERS			DOCTORATE		
PROVINCES	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX	CV=16.5 % MINX	CV=25 % MINX	CV=33.3 % MINX
CANADA	704	308	174	176	77	44	70	31	18
NEWFOUNDLAND	135	61	35	46	23	14	-	-	-
P.E.I.	53	25	14	-	-	-	-	-	-
NOVA SCOTIA	254	113	65	66	30	17	-	-	-
NEW BRUNSWICK	147	66	38	47	22	13	-	-	-
QUEBEC	571	252	143	175	78	44	63	29	17
ONTARIO	1,098	484	274	214	95	54	67	30	17
MANITOBA	233	104	59	49	22	13	-	-	-
SASKATCHEWAN	229	102	58	48	22	13	-	-	-
ALBERTA	370	165	94	81	37	21	62	31	18
BRITISH COLUMBIA	432	193	110	108	49	28	64	32	19

10.6 C.V. Tables

Please refer to the files identified below for the C.V. tables for the Follow-up of 1990 Graduates micro data:

CVTABENG.WP6 or CVTABENG.PDF

CVTABFRE.WP6 or CVTABFRE.PDF

11.0 Weighting

This chapter outlines the notation and formulae used in producing population estimates for characteristics of graduates as well as for the calculation of sampling variabilities for these estimates. The respondents to the Survey of 1990 Graduates were selected for the Follow-up Survey. Therefore, the basic procedure to calculate weights for the Follow-up Survey is the same as that of the Survey of 1990 Graduates, with adjustments for non-response. Further details are provided below:

Notation

I: subscript denoting I th responding record (I=1,...,n)

p: subscript denoting p th province (of institution)

l: subscript denoting l th level of certification as reported by the institution (l=1,...,5)

m: subscript denoting the m th major field of study as reported by the institution

(m=1,...9, for university and career/technical programs)
(m=1,...,10 for trade/vocational programs)

N: total number of graduates in population as reported by institutions in 1992.

N_{plm}: number of graduates in (p, l, m) as reported by institutions in 1992 (note that this excludes 1990 Québec graduates with diplomas or certificates below and above the Bachelor level).

n'd: number of graduates selected for sample who responded to survey from domain d in 1995.

W_i^{92}: 1992 final sampling weight for persons selected from (p, l, m)

W_i: 1995 final sampling weight for persons selected from (p, l, m)

X_{di}: value of characteristic X taken on by I th responding record in domain d.

$\hat{X}_d$: estimated total number of graduates with characteristic X in domain d.

$\hat{\bar{X}}_d$: estimated mean number of graduates with characteristic X in domain d.

Adj_1: first non-response adjustment for persons selected from (p, l, m) in 1995

Adj_2: second non-response adjustment for persons selected from (p, l, m) in 1995

Adj_3: third non-response adjustment for persons selected from (p, l, m) in 1995

$\hat{N}_d$: estimated size (number of persons) of domain d (e.g., persons still living in Canada).

$\hat{P}_d$: estimated proportion of units (having characteristic X) in domain d.

$\mathrm{V}\hat{a}r$: estimated variance of an estimate.

Weights:

As a result of the Follow-up Survey, 1990 graduates were classified into one of the seven following categories:

A: refers to graduates who gave complete or partial information to the Follow-up Survey and agreed to share their data (see item I27 of the Follow-up Survey questionnaire);

B: refers to graduates who were no longer living in Canada at the time of the Follow-up Survey;

C: refers to graduates we contacted but who refused to participate in the Follow-up Survey or who were unable to be interviewed due to other reasons (illness, language problems, etc.);

D: refers to graduates who were absent for the duration of the Follow-up Survey, had no listed telephone number, could not be reached by phone or did not answer the phone after several tries;

E: refers to graduates we were unable to trace;

F: refers to graduates who responded to the 1992 survey but were deleted from the Follow-up sample in 1995 due to misclassification of their stratum identifiers during processing in 1992;

G: refers to graduates who refused to share their data (see item I27 of the Follow-up Survey questionnaire), or graduates who were re-classified from respondents to non-respondents during 1995 processing due to missing/insufficient data.

From the notation on the previous pages

$\mathrm{Adj}_1 = \dfrac{\sum_{ieCDG} W_i^{92}}{\sum_{ieA} W_i^{92}}$ if respondent I belongs to category A

$\mathrm{Adj}_1 = 0$ otherwise

$\mathrm{Adj}_2 = \dfrac{\sum_{ieEF} W_i^{92}}{\sum_{ieAB} W_i^{92}}$ if respondent I belongs to either category A or B

$\mathrm{Adj}_2 = 0$ otherwise

$\mathrm{Adj}_3 = \&1$ if respondent I belongs to either categories C, D, E, F, or G

$\mathrm{Adj}_3 = 0$ otherwise

$W_i = W_i^{92} \; x \; (1 \% \mathrm{Adj}_1 \% \mathrm{Adj}_2 \% \mathrm{Adj}_3)$

where $\sum_{ieAB} W_i^{92}$ means the sum of the 1992 weights W_i^{92} from (p, l, m) for respondents who belong to categories A or B

Note that a respondent to the Follow-up Survey must be in one and only one of the categories A to G defined above.

We are now in a position to define one more notation:

n_{plm}^{95}: refers to the number of graduates who belong to categories A or B.

It is important to note that records which were classified in category B were treated differently from other kinds of non-responses (namely categories C, D, E, F, and G) in the determination of the sampling weights.

This was because the Follow-up Survey frame contains not only the target population, but also graduates no longer living in Canada.

Records which belong to category B, being outside the target population, were excluded from files used for tabulation and analysis but these graduates must retain their weights to account for the contamination of the frame.

ESTIMATES

Estimates of totals for attribute variables (e.g. number of graduates enrolled in a one-year course) must be derived by summing the final weight of all records possessing the attribute. Estimates for quantitative (or continuous) variables (e.g. age of graduate) must be derived by multiplying the value of the variable by the final weight and summing this product over all records, then dividing by the sum of the weights if a mean is required.

The formulae following were applicable for deriving these types of estimates:

$\hat{X}d = \sum_{i=1}^{n)d} W_i \, x_{di}$ Where x_{di} = 1 if unit I belongs to domain of interest d and has characteristic x

= 0 otherwise

$\hat{N}d = \sum_{i=1}^{n)d} W_i$

$\hat{P}d = \frac{\hat{X}d}{\hat{N}_d} = \frac{\sum_{i=1}^{n)d} W_i \, x_{di}}{\sum_{i=1}^{n)d} W_i}$ Where x_{di} = 1 if unit I belongs to domain of interest d and has characteristic x

= 0 otherwise

$\hat{\bar{Y}}d = \frac{\hat{Y}d}{\hat{N}d} = \frac{\sum_{i=1}^{n)d} W_i \, y_{di}}{\sum_{i=1}^{n)d} W_i}$ Where y is a continuous variable (for example, gross income) and $y_{di} = y_i$ if unit I is in domain of interest d.

SAMPLING VARIABILITIES

Estimates of the variance of a total estimated from this survey can be calculated using the Approximate Sample Variability Tables or the following formula:

$$\hat{Var}_{st}(\hat{Yd}) = \sum_{p}\sum_{l}\sum_{m}\left[N_{plm}\frac{(N_{plm}-n_{plm}^{95})}{n_{plm}^{95}}\ \frac{\sum_{i}(y_{plmd_i}-\bar{y}_{plm})^2}{(n_{plm}^{95}-1)}\right]$$

Where $y_{plmd_i} = y_{plm_i}$ if unit I is in the domain of interest d

(This is a usual variance estimator for a stratified survey with simple random sampling within a stratum.)

P.S. n_{d} can be used instead of n_{plm}^{95} since the latter is not available from the files used for tabulations.

12.0 Questionnaires and Code Sheets

12.1 Survey of 1990 Graduates Questions

Please refer to the files identified below for the Survey of 1990 Graduates questions:

S90GQUEN.WP6 OR S90GQUEN.PDF

S90GQUFR.WP6 OR S90GQUFR.PDF

12.2 Follow-up of 1990 Graduates Questions

Please refer to the files identified below for the Follow-up of 1990 Graduates questions:

F90GQUEN.WP6 OR F90GQUEN.PDF

F90GQUFR.WP6 OR F90GQUFR.PDF

13.0
Record Layout and Univariates

Please refer to the files identified below for the codebooks for the Survey of 1990 Graduates and the Follow-up of 1990 Graduates micro data:

G90ECBKE.WP6 OR G90ECBKE.PDF

G90ECBKF.WP6 OR G90ECBKF.PDF