Microdata User Guide

2013 National Graduate Survey (Class of 2009-2010)

Public Use Microdata File

Statistique Canada
Statistics Canada

Canada

Table of Contents

1.0 Introduction

The 2013 National Graduates Survey – Class of 2009/2010 was conducted by Statistics Canada from April 2nd to September 1st, 2013. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

The public use microdata file, or PUMF, contains a reduced list of variables compared to the Master file. The need to preserve the confidentiality of respondents dictated that many variables that could have been used to identify individuals be removed from the file. In addition, all continuous variables such as those relating to income, student loans or age at graduation, were converted to categorical variables, and many existing categorical variables were grouped into a smaller number of categories. Finally, local suppression was used where necessary to further protect confidentiality. Every effort was made to preserve the analytical utility of the data during this process.

It is also important to note that the PUMF contains fewer records than the Master file. As an initial measure of diminishing the risk of disclosure, a subsample of the records from the Master file was drawn. The PUMF therefore is made up of 14,745 records, or roughly half the number in the Master. Users should be aware that estimates produced using the subsample may not correspond exactly to those produced by Statistics Canada using the Master file.

Users requiring access to information excluded from the microdata files may purchase custom tabulations.

This document retains most of the content from the original user guide for the NGS master microdata file for informational purposes. Notes have been added to indicate where full content is not applicable to the PUMF.

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

Statistics Canada

Client Services
Centre for Education Statistics
Room SC-2000 B, Main Building
150 Tunney's Pasture Driveway
Ottawa, Ontario
K1A 0T6

Telephone: (613) 951-7608 or call toll-free 1 800 307-3382
Fax: (613) 951-9040
E-mail: educationstats@statcan.gc.ca

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. 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 has 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.

Graduation Year	Survey Year	Survey Name
1976	1978	Survey of 1976 Graduates of Post-Secondary Programs
1982	1984	Survey of 1982 Graduates (S82G) (also known as the National Graduates Survey or NGS)
	1987	Follow-up of 1982 Graduates (F82G) (also known as the Follow up of Graduates or FOG)
	1984 & 1987	PUMF – Survey of 1982 Graduates and Follow-Up of 1982 Graduates combined
1986	1988	Survey of 1986 Graduates (S86G)
	1988	PUMF – Survey of 1986 Graduates
	1991	Follow-up of 1986 Graduates (F86G)
1990	1992	Survey of 1990 Graduates (S90G)
	1992	PUMF – Survey of 1990 Graduates
	1995	Follow-up of 1990 Graduates (F90G)
1995	1997	Survey of 1995 Graduates (S95G)
	1997	PUMF – Survey of 1995 Graduates
	2000	Follow-up of 1995 Graduates (F95G)
2000	2002	National Graduates Survey - Class of 2000 (NGS2000)
	2005	Follow-up of Graduates Survey - Class of 2000 (FOG2000)
	2002 & 2005	PUMF – National Graduates Survey (2000) and Follow-Up of Graduates (2000) combined
2005	2007	National Graduates Survey - Class of 2005 (NGS2005)

	2007	PUMF – National Graduates Survey (2005)
2009/2010	2013	National Graduates Survey – Class of 2009/2010 (NGS2013)
	2013	PUMF – National Graduates Survey (2009/2010)

The survey contains data on: the link between education experience and labour market outcomes; information regarding the job held in the week prior to the interview and the first job after graduation; financial and loan information; additional education pursued after graduation; and socio-economic background.

In comparison to the NGS2005 questionnaire, the main changes that were made to the NGS2013 questionnaire are the following:

Questions were added or reworded to match harmonized content (i.e. a process developed to ensure standard concepts, definitions, classification and wording across all household surveys).

New concepts were added on lifelong learning, mobility, skills and qualifications mismatch, intended occupation at graduation, student loans, debt management and sources of support during graduate school.

Note, however, that for confidentiality reasons, information specific to graduates who took trade / vocational programs, graduates who lived in the United States, and information on components of programs taken outside of Canada is not available on the PUMF.

3.0 Objectives

The survey's primary objective is to obtain information on the labour market experiences of graduates entering the labour market, focusing on employment, occupations and the relationship between jobs and education.

The survey's key data objectives are:

- To obtain information for labour market analysis of a key youth group at an important time, focusing on education, training, employment, occupations and geographic mobility. The data and analysis will be useful for policy development.
- To obtain information on the exposure of graduates to additional learning opportunities.
- To extend available information required to improve occupational supply and demand projection models for various occupational categories.
- To obtain data regarding longer-term labour market experiences of graduates, with special emphasis on employment and occupations, for use in counselling on career and post-secondary education course selection.
- To obtain information on labour market experiences of members of target groups (such as women, native people and the disabled), which permits longitudinal and comparative analysis useful in the formulation of job equity policies.
- To gain a better understanding of school-work transitions and returns to human capital.
- To gain a better understanding of post-secondary education financing.
- To obtain more detailed information on knowledge and skills.

4.0 Content

The following table describes the content of each section of the National Graduates Survey Class of 2009/2010 (NGS2013) questionnaire.

Section	Content
Program confirmation (PR)	Graduates are asked to give information about the program they graduated from in 2009/2010.
Activities before graduation (AB)	Contains information about the graduate's activities (i.e. employment, education, etc.) prior to graduating in 2009/2010.
Graduates who live/lived in the United States (US and MU)	Identifies graduates who moved to the United States (US) after graduation and obtains information on their activities in the US and also about their return to Canada, if applicable. These variables are not available on the PUMF.
Activities last week (LF)	Asks about the graduate's labour force activity the week before the interview.
First job after graduation (EM)	Contains information about the first job held by the graduate after graduation.
First job in the United States (FU)	Contains information about the first job in the United States, if applicable and not already collected in sections LF or EM. These variables are not available on the PUMF.
Education programs (ED) Education program description (EP)	Asks about completed and uncompleted educational programs taken after graduation.
Student loans (ST)	Asks questions about student loans and finances.
Higher education (HE)	Contains information about the intentions of graduates to pursue a Master's degree or a Ph.D.
Demographic characteristics (DE and DEM)	Asks general questions such as marital status, number of dependent children, income, and disabilities.

4.1 Concepts and Definitions

Graduation date

For the purpose of this survey, the graduation date is the year and month in which the graduate completed the requirements (PR_Q11A and PR_Q11B) of his/her program. To complete the requirements of the program, graduates must have written and passed the last exam, submitted the last paper, report or project for a program or defended a thesis. These variables are not available on the PUMF.

Graduates who moved to live in the United States

Graduates who live in the United States, or lived in the United States since their graduation but have returned to Canada, are included in the survey. They may have moved to attend school, to work or to accompany a partner or spouse. Anyone who visited or vacationed in the United States temporarily is not considered to have moved.

These variables are not available on the PUMF.

Transition after completing post-secondary studies

A number of modules in the survey are devoted to obtaining information on the graduate's activities after completing his/her post-secondary studies. The information found in these modules allows for a detailed analysis on the graduate's transition after completing his/her post-secondary studies.

- The LF module asks about the graduate's labour force activities during the week prior to the interview (i.e., employed, unemployed, or not in the labour force). Detailed information on the job held in the week prior to the interview is also collected.

- The EM module obtains information about the first employer the graduate worked for after graduation, and detailed information about the job held with this employer (or equivalent information if the respondent was self-employed).

- The ED and EP modules collect information on completed and uncompleted educational programs taken after graduation when these programs lead towards a diploma, certificate or degree that would take someone three months or more to complete if taken full-time.

Main job

The job involving the greatest number of hours per week.

Paid worker (LMA3_Q10)

A person who works for others (i.e. works for an employer). Payment may be in cash (salary, wages, tips, commissions) or "payment in kind" (payment in goods or services rather than money). Such employer-employee relationships almost always involve some legal obligations on the part of the employer, must deduct and remit income tax and Canada/Québec Pension Plan premiums, etc.

Self-employed (LMA3_Q10)

A person who works directly for himself/herself. A self-employed person may or may not have a business, farm or professional practice. Examples of self-employed persons with a business would be: a man with his own barber shop, or a woman with her own medical practice.

Examples of self-employed persons without a business include:
- a cleaning person working for a number of people in their homes;
- a freelance writer, a paper carrier;
- a general handyman;
- a caregiver who works for a number of people.

Self-employed and unpaid family workers have been combined on the PUMF.

Unpaid family worker (LMA3_Q10)

An unpaid family worker is someone who worked without pay on a farm or in a business owned and operated by another family member living in the same household. The work done must contribute directly to the operation of a family farm or family business.

Self-employed and unpaid family workers have been combined on the PUMF.

Permanent job (LF_Q24)

A permanent job is one that is expected to last as long as the employee wants it and as long as business conditions permit. That is, the employer did not hire the employee on the understanding the job would end at a specified time in the near future. Sometimes permanent jobs are referred to as indeterminate, since they have no pre-specified date of termination.

Non-permanent job (LF_Q24)

A job that is not permanent is one that has a predetermined date on which it will end or will end as soon as a specified project is completed. The employer has hired the employee on the understanding that the job will end at this specified time in the near future.

Seasonal job (LF_Q25)

This occurs in industries where employment levels rise and fall with the seasons (seasonal employment).

Examples: farming, fishing, logging and the tourist industry.

This variable is not available on the PUMF.

Temporary, term or contract job (non-seasonal) (LF_Q25)

A job in which there was a definite indication from the employer before the job was accepted that the job would terminate at a specified point in time, or at the end of a particular task or project.

This variable is not available on the PUMF.

Casual job (LF_Q25)

Is one of the following:

- respondent has work hours that vary substantially from one week to the next;
- respondent is called to work by the employer when the need arises, not on a pre-arranged schedule; or
- respondent does not usually get paid for time not worked and there is no indication from the employer that he/she will be called to work on a regular, long-standing basis.

This variable is not available on the PUMF.

Number of (paid) hours worked per week (LMA6_Q16)

Serves to separate the employed into full-time (30 hours of work or more per week) and part-time (less than 30 hours of work per week) workers.

Number of paid hours usually worked is asked of employees.

Number of hours usually worked is asked of self-employed persons.

This variable is not available on the PUMF.

Wages or salary (LF_Q83)

For employees, this refers to wages before deductions by the employer for taxes, employment insurance (EI), government pension plans (CPP/QPP), union dues, etc. (referred to as "other deductions"). Most pay cheques are received weekly or every two weeks but some respondents only know their salaries/wages before taxes and deductions on a monthly or annual basis. The respondent may choose any reporting period, which makes it easier for him/her to give accurate data.

This variable is not available on the PUMF.

Tips and commissions

Tips, bonuses or commissions are averaged over the period for which they apply and included with the wages or salary reported. This applies to weekly, bi-weekly, semi-monthly, monthly and yearly wages.

Government sponsored student loan

A loan sponsored by the federal government or any provincial/territorial government, which enables the respondent to finance their studies.

As of March 2001, Canada Student Loans come directly from the Government of Canada through the National Student Loans Service Centre. The loan is either deposited or mailed to the individual.

From August 1, 1995 up to March 2001, Canada Student Loans were issued by banks, Credit Unions and Caisses Populaires but were guaranteed by the government.

"Student loan" applies to any education, not just the program from which the respondent graduated. It could include undergraduate and graduate programs.

Scholarships, awards, fellowships, prizes

Merit-based (i.e. based on individual achievements) financial assistance to help students continue their studies. These may be awarded by governments or by private donors. Scholarships, awards, fellowships and prizes apply to any education, not just the program from which the respondent graduated. It could include undergraduate and graduate education.

Grants, bursaries

Financial assistance to students which is need-based and/or targeted for specific purposes.

A grant is a gift (usually a sum of money) made by a government or corporation (as an educational or charitable foundation) to a beneficiary on the condition that certain terms be accepted or certain engagements fulfilled which are required by the sponsor.

A bursary refers to a monetary award to assist a student in the pursuit of his/her studies based on financial need and satisfactory achievement.

Grants and bursaries apply to any education, not just the program from which the respondent graduated. It could include undergraduate and graduate education.

Income

The income information is for the income received from all sources by the graduate in the calendar year 2012. It is not limited to monies that are taxable.

It includes:

- income from wages and salaries;
- net income from self-employment;
- regular Employment Insurance benefits as well as those for sickness, maternity or paternity leave, adoption, job creation, work sharing, retraining and benefits to self-employed fisherman;
- retraining and retirement benefits received under the Employment and Social Development Canada (formerly Human Resources and Social Development Canada) employment insurance program;
- payments from provincial or municipal programs for persons in need such as Social Assistance or welfare;
- spousal support or child support;
- scholarships, grants, bursaries or fellowships;
- money from the Canada or Quebec Pension Plan;
- Canada Child Tax Benefits or provincial child tax benefits or credits;
- interest from Canadian and foreign sources;
- foreign dividends;
- taxable dividends received from Canadian corporations;
- net rental income;
- rents for leased farm land;
- regular income from an estate or trust fund;
- cash dividends from life insurance policies;
- pensions from deferred profit sharing plans and other private pension plans; and
- money from parents, guardians or others that does not have to be repaid.

It excludes:

- monies received from student loans or any other loan;
- income tax refunds;
- tax-free Registered Retirement Savings Plan withdrawals used for purchasing a home;

- proceeds from the sale of property, businesses, financial assets or personal belongings;
- loans repaid to the graduate as a lender; and
- refund of contributions to work-related pension plans.

4.2 Uses

Following from previous surveys, this survey extends the existing base of information on the labour-market experiences of recent graduates. 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 is applied.

- The survey data can be used 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.
- Job equity programs will receive important labour market related information on designated groups such as women, aboriginal peoples, persons with disabilities and visible minorities.
- The survey provides concrete information regarding graduates' labour market experiences during the two years after graduation. This information can be used to aid post-secondary education course selection and career counselling.

5.0 Survey Methodology

The National Graduates Survey – Class of 2009/2010 (NGS2013) is a cross-sectional survey designed to collect data from Canadian graduates**.**

5.1 Target Population

The target population of the NGS2013 consists of all graduates from a recognized public post-secondary Canadian institution who completed the requirements of an admissible program or obtained a diploma some time in 2009/2010, and who were living in Canada or the United States at the time of the survey (with the exception of American citizens living in the United States at the time of the survey).

These graduates include:

- graduates of university programs that lead to bachelor's, master's or doctoral degrees, or that lead to specialized certificates or diplomas;

- graduates of post-secondary programs (that is, programs that normally require a secondary school completion or its equivalent for admission) in Colleges of Applied Arts and Technology (CAAT), Collèges d'enseignement général et professionnel (CEGEP in Quebec), community colleges, technical schools or similar institutions; and

- graduates of skilled trades (that is, pre-employment programs that are normally three months or more in duration). A trade/vocational school is a public educational institution that offers courses to prepare people for employment in a specific occupation such as heavy equipment operator, automotive mechanic or upholsterer. Many community colleges and technical institutes offer certificates or diplomas at the trade level.

The survey excludes:

- graduates from private post-secondary institutions (for example, computer training and commercial secretarial schools);

- graduates who completed "continuing education" courses at universities and colleges (unless they led to a degree or diploma); and

- graduates in apprenticeship programs.

5.2 Survey Frame

The survey frame for the 2009/2010 graduates was created by Statistics Canada’s Centre for Education Statistics from a list of all graduates from the Post-Secondary Information System (PSIS), universities, colleges and trade/vocational schools in Canada.

Data on graduates were provided through two sources: the main source of information was from the individual institutions and provincial co-ordinating bodies, while the second source of graduate data came from the Postsecondary Student Information System (PSIS), which is maintained by the Centre for Education Statistics.

Where the PSIS data could not be extracted, files of graduates, preferably in machine-readable form, were requested from the institutions or provincial co-ordinating bodies. The same information that is submitted to the PSIS was requested for each graduate: his/her name,

permanent address and telephone number, local address and telephone number, qualification obtained in 2009/2010, major field of study, date of birth, student number, immigration status, gender, mother tongue, graduation date and whether the program taken was a co-op program.

5.2.1 CIP coding

A standard Classification of Instructional Programs (CIP 2000) code was assigned to all graduates on the frame. This coding process is mostly automated as it is already a regular process for PSIS, but some of the cases were coded manually. The CIP code was required to derive the field of study variable used for stratification. It was also used to eliminate from the frame graduates from programs that are not part of the target population.

A derived variable (PRCIPAGP) was created for the PUMF.

5.2.2 Removal of duplicates

A verification of duplicates was done on the survey frame. Duplicates consist of two or more records on the frame that refer to the same person and that are classified in the same stratum (see Section 5.3.2 for the stratum definition). When duplicates were found, only one record was kept on the survey frame for that person. Note that when a person graduated in two different programs (programs falling into two different strata), both records of this person were kept on the survey frame. However, if both records were selected in the sample, that person was contacted only once.

5.3 Stratification

The NGS2013 uses a stratified simple random sample design. The sample selection of graduates within strata is done without replacement and using a systematic method.

Three variables are used for stratification; geographical location of the institution, level of certification and field of study. There are 13 geographical locations: the ten provinces and the three northern territories. There are 5 levels of certification: trade/vocational certificate or diploma, college diploma, bachelor's degree, master's degree, and doctorate. The "trade/vocational" level only exists in Quebec; in all other geographical locations, these units are part of the second group (college). Finally, there are 12 fields of study: categories 010 to 120 of the primary groupings of the Classification of Instructional Programs (CIP 2000). Details about the field of study can be found in Appendix A1. The combination of these three variables makes for a possibility of 636 strata in total. However, there are not graduates in every possible strata and therefore, the final number of strata created was 434.

5.4 Sample Allocation, Selection and Size

The sample is designed to yield estimates of a minimal proportion of 5.5% with a maximum coefficient of variation (CV) of 16.17% for any of the NGS2013's marginal. A marginal is defined as: i) a given field of study regardless of the province of institution or ii) a given province of institution regardless of the field of study; and that for each of the five levels of certification. There were two exceptions to this rule: the sample sizes for "Other" fields of study were reduced by two thirds (since they aren't as important from an analytical point of view), and all PhDs were drawn into the sample.

The marginal's CVs are then allocated to each stratum (or cell in a table) to obtain the cells or stratums CV using a raking-ratio algorithm. The last step consists of converting the CV's into sample sizes.

Note that the expected non-response and out-of-scope rates were taken into account when establishing the sample sizes. A few units, despite being drawn into the original sample, were not sent to collection; they were automatically coded as non-respondents because, for example, the respondent's name was not on the frame.

The table below presents the distribution of the population and the sample sizes that were sent to collection, by province/territory and level of certification. The population sizes represent the number of graduates on the initial frame.

Frame and Sample Size by Province / Territory and Level of Certification

Province / Territory by Level of Certification	Number of Units on Frame	Sample Size Sent to Collection
Newfoundland and Labrador	**6,015**	**2,966**
College diploma	2,613	1,090
Bachelor's degree	2,781	1,362
Master's degree	559	452
Doctorate	62	62
Prince Edward Island	**2,012**	**1,278**
College diploma	1,315	667
Bachelor's degree	645	564
Master's degree	42	37
Doctorate	10	10
Nova Scotia	**13,773**	**4,503**
College diploma	4,284	1,349
Bachelor's degree	7,570	1,934
Master's degree	1,814	1,115
Doctorate	105	105
New Brunswick	**7,246**	**3,367**
College diploma	2,375	1,352
Bachelor's degree	4,224	1,468
Master's degree	597	497
Doctorate	50	50
Quebec	**139,824**	**15,276**
Trade/vocational	44,905	5,810
College diploma	28,343	2,367
Bachelor's degree	49,637	2,968
Master's degree	15,216	2,574
Doctorate	1,723	1,557
Ontario	**161,633**	**9,372**
College diploma	63,981	2,329
Bachelor's degree	79,064	2,402
Master's degree	16,360	2,422
Doctorate	2,228	2,219
Manitoba	**11,434**	**3,663**
College diploma	3,918	1,486
Bachelor's degree	6,709	1,778
Master's degree	695	287
Doctorate	112	112

Province / Territory by Level of Certification	Number of Units on Frame	Sample Size Sent to Collection
Saskatchewan	**22,844**	**3,464**
College diploma	16,351	913
Bachelor's degree	5,355	1,737
Master's degree	994	670
Doctorate	144	144
Alberta	**37,156**	**7,504**
College diploma	15,677	3,101
Bachelor's degree	17,505	2,284
Master's degree	3,363	1,513
Doctorate	611	606
British Columbia	**53,419**	**7,625**
College diploma	24,162	2,593
Bachelor's degree	24,070	2,501
Master's degree	4,533	1,877
Doctorate	654	654
Yukon	**146**	**143**
College diploma	109	107
Bachelor's degree	37	36
Northwest Territories	**183**	**181**
College diploma	182	180
Bachelor's degree	1	1
Nunavut	**125**	**125**
College diploma	125	125
Canada	**455,810**	**59,467**
Trade/vocational	44,905	5,810
College diploma	163,435	17,659
Bachelor's degree	197,598	19,035
Master's degree	44,173	11,444
Doctorate	5,699	5,519

6.0 Data Collection

Project supervisors and Senior interviewers from the Statistics Canada Regional Offices came to head office for a one-day classroom training seminar. Presentations on subject matter and methodology were made, along with mock interviews. Project supervisors and Senior interviewers then conducted a one-day training of interviewers in the Regional Offices, assisted with an interactive tutorial and mock interviews.

Interviewers collected the data using a computer-assisted telephone interviewing method (CATI). They were instructed to make all reasonable attempts to obtain interviews with the selected graduates. Proxy response was not allowed. For graduates who refused to participate, a letter was sent from the Regional Office to the dwelling address stressing the importance of the survey and the graduate's cooperation. This was followed by a second call from the interviewer. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there was no one home, numerous call backs were made. If graduates had moved, various tracing methods were used to locate them.

The collection period was scheduled to run from April 2nd to September 1st 2013.

7.0 Data Processing

This chapter presents a brief summary of the processing steps involved in producing the microdata file.

7.1 Data Capture

Responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors, and data transmission. The response data are transmitted over a secure line to Ottawa.

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

7.2 Editing

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

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

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

The third type of editing performed was related to inconsistencies in some of the responses received. In a situation where an inconsistency was found, depending on the nature of the inconsistency, various actions could be taken. The inconsistent variable (or one of the variables involved) could either be changed to “not stated”, corrected or left unchanged. For example If a respondent reported an hourly salary of 35,000 dollars, the “hourly” was changed to “annually”. However, in situations where it was not possible to determine which variable was most likely to be wrong, no action was taken and a flag was derived.

For quantitative variables such as financial variables, editing which includes outlier detection was performed. These variables include reported information on earnings, income, and student loans. Potential outliers were identified and manual investigations were made on these cases to confirm their outlier status. Outliers were changed to “not stated” or replaced by a more plausible value when a realistic value could be deduced from the other variables.

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. These were items relating to the type of education programs taken before and after graduation in 2009/2010, as well as questions relating to the graduates’ industry and occupation. These open-ended questions were coded using various standard classifications (see Sections 7.3.1 and 7.3.2). An additional type of coding performed is called “Other – Specify” coding (see Section

7.3.3).

7.3.1 Coding of Education Programs

Field of study program descriptions were coded using the Classification of Instructional Programs (CIP 2000). Programs were coded at the six-digit level. See Appendix A1 for details on the code set. Field of study programs in module PR were also coded using the CIP 2011 (see Appendix A2).

A derived variables (PRCIPAGP) was created for the PUMF. See Appendix A1.

7.3.2 Coding of Industry and Occupation

For each job held by the graduate 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 graduate in the job. This information was used to assign industry and occupation codes to each job using the North American Industry Classification System (NAICS) 2012 and the National Occupational Classification (NOC) 2011. See Appendix B and C for details on the code sets. For the user's convenience, the NAICS and the NOC variables have been grouped in their own section in the codebook.

A derived variable (LFCINDP) was created for the PUMF. See Appendix B.
A derived variable (LFCOCCP) was created for the PUMF. See Appendix C.

7.3.3 Coding of "Other – Specify" Answers

"Other – Specify" coding was done on questions that contained a list of answer categories that had "Other - Specify" as the final category. If the write-in was reflected in one of the existing categories, the response was recoded into the appropriate one. Responses that could not be coded into an existing category or into new categories were coded as "Other".

7.4 Imputation

No imputation was done for the National Graduates Survey – Class of 2009/2010 (NGS2013).

7.5 Creation of Derived Variables

Combining Items

A number of variables have been derived by combining questions on the questionnaire in order to facilitate data analysis. For example, questions from the Activities Last Week section (i.e. module LF, LMAM and LMA2) section are used to derive labour force status in the week prior to the interview (LFSTAT). These included:

- LF_Q01 - Did you attend school, college, CEGEP or university (last week)?
- LF_Q02 - Were you enrolled as...?

- LMAM_Q01 - Many of the following questions concern (your) activities last week. By last week, I mean the week beginning on (Monday) and ending (Sunday). Last week, did (you) work at a job or business? (regardless of the number of hours)
- LMAM_Q03 - What was the main reason (you were) absent from work last week?
- LMA2_Q04 - In the 4 weeks ending (Sunday), did (you) do anything to find work?
- LMA2_Q05 - Last week, did (you) have a job to start at a definite date in the future?
- LMA2_Q06 - Will (you) start that job before or after (Monday)?
- LMA2_Q07 - Did (you) want a job with more or less than 30 hours per week?
- LMA2_Q09 - What was the main reason (you were) not available to work last week?
- AGEINT - Derived Variable- Respondent's age at interview.

Where to find the Derived Variables on the File

All the derived variables have been grouped at the end of the codebook by module. Within each module, the derived variables have been arranged in alphabetical order. For a complete list of the derived variables on the Master file and on the PUMF, and a description on how they were derived, see Appendix H.

8.0 Response Rates

This chapter describes the response rates for NGS2013. Survey response rates are measures of the effectiveness of the population being sampled and the collection process. They are also a good indicator of the quality of the estimates produced.

In-scope records are records that met all criteria in the target population as defined in Section 5.1. A respondent is a person for whom there is usable minimal information on the questionnaire. Cases where the graduates did not go far enough in the questionnaire – defined as not having answered the "first job after graduation" module –were deemed non-responding units.

Table 8.1 presents the collection results for NGS2013. The first column is the sample size sent to collection, while the second column is the ones that remain after removing the units that were found to be out of scope during collection. The "Realized Number of Responding Graduates" gives the final number of respondents, and unlike the first two columns, it reflects the Province/Territory and Level of Certification that was reported during the interview – which wasn't necessarily the same as the one on the frame. Therefore, one must be careful while interpreting the response rates. For example: in Newfoundland and Labrador, the Master file includes 614 College respondents out of 1,064. However, the 614 are not necessarily a subset of the 1,064: some of the 614 may have initially been thought to be Bachelor's degrees in Newfoundland, so they would in fact come from that group of 1,347. Similarly, the table suggests that there were 450 non-responding units from Newfoundland and Labrador Colleges (1,064 – 614 = 450), but some of these may have in fact become respondents for a different level of certification (or in a different province or territory, but such cases were rare). More details on the non-response adjustments are provided in section 9.

The following two types of response rates are presented in the table:

Response Rate – Master File =

$$\frac{\text{Number of responding graduates on Master File}}{\text{Number of in-scope graduates}}$$

Response Rate – Share File =

$$\frac{\text{Number of responding graduates who agreed to share their data}}{\text{Number of in-scope graduates}}$$

Table 8.1 Response Rate by Province / Territory and Level of Certification – Unweighted

Province / Territory by Level of Certification	Sample Size Sent to Collection	In-scope Sample	Realized Number of Responding Graduates		Response Rate (%)	
			Master	Share	Master	Share
Newfoundland and Labrador	**2,966**	**2,915**	**1,651**	**1,569**	**56.6**	**53.8**
College diploma	1,090	1,064	614	579	57.7	54.4
Bachelor's degree	1,362	1,347	736	704	54.6	52.3
Master's degree	452	443	279	267	63.0	60.3
Doctorate	62	61	22	19	36.1	31.1
Prince Edward Island	**1,278**	**1,272**	**634**	**605**	**49.8**	**47.6**
College diploma	667	665	305	293	45.9	44.1

Province / Territory by Level of Certification	Sample Size Sent to Collection	In-scope Sample	Realized Number of Responding Graduates		Response Rate (%)	
			Master	Share	Master	Share
Bachelor's degree	564	560	298	282	53.2	50.4
Master's degree	37	37	25	24	67.6	64.9
Doctorate	10	10	6	6	60.0	60.0
Nova Scotia	**4,503**	**4,469**	**2,610**	**2,520**	**58.4**	**56.4**
College diploma	1,349	1,339	851	812	63.6	60.6
Bachelor's degree	1,934	1,922	1,007	974	52.4	50.7
Master's degree	1,115	1,104	701	683	63.5	61.9
Doctorate	105	104	51	51	49.0	49.0
New Brunswick	**3,367**	**2,986**	**1,579**	**1,472**	**52.9**	**49.3**
College diploma	1,352	989	452	400	45.7	40.4
Bachelor's degree	1,468	1,462	824	780	56.4	53.4
Master's degree	497	485	276	265	56.9	54.6
Doctorate	50	50	27	27	54.0	54.0
Québec	**15,276**	**15,140**	**6,868**	**6,519**	**45.4**	**43.1**
Trade/vocational	5,810	5,716	2,550	2,372	44.6	41.5
College diploma	2,367	2,348	1,459	1,391	62.1	59.2
Bachelor's degree	2,968	2,954	1,285	1,241	43.5	42.0
Master's degree	2,574	2,566	903	868	35.2	33.8
Doctorate	1,557	1,556	671	647	43.1	41.6
Ontario	**9,372**	**9,294**	**4,659**	**4,391**	**50.1**	**47.2**
College diploma	2,329	2,308	1,167	1,076	50.6	46.6
Bachelor's degree	2,402	2,383	1,226	1,152	51.4	48.3
Master's degree	2,422	2,408	1,320	1,266	54.8	52.6
Doctorate	2,219	2,195	946	897	43.1	40.9
Manitoba	**3,663**	**3,628**	**1,941**	**1,786**	**53.5**	**49.2**
College diploma	1,486	1,470	773	696	52.6	47.3
Bachelor's degree	1,778	1,764	922	858	52.3	48.6
Master's degree	287	283	178	169	62.9	59.7
Doctorate	112	111	68	63	61.3	56.8
Saskatchewan	**3,464**	**3,399**	**1,771**	**1,679**	**52.1**	**49.4**
College diploma	913	868	441	409	50.8	47.1
Bachelor's degree	1,737	1,724	908	878	52.7	50.9
Master's degree	670	663	353	332	53.2	50.1
Doctorate	144	144	69	60	47.9	41.7
Alberta	**7,504**	**7,395**	**3,671**	**3,496**	**49.6**	**47.3**
College diploma	3,101	3,034	1,482	1,390	48.8	45.8
Bachelor's degree	2,284	2,261	1,123	1,079	49.7	47.7
Master's degree	1,513	1,496	771	746	51.5	49.9
Doctorate	606	604	295	281	48.8	46.5
British Columbia	**7,625**	**7,533**	**3,179**	**2,986**	**42.2**	**39.6**
College diploma	2,593	2,545	1,027	944	40.4	37.1
Bachelor's degree	2,501	2,482	1,083	1,023	43.6	41.2
Master's degree	1,877	1,857	824	783	44.4	42.2

Province / Territory by Level of Certification	Sample Size Sent to Collection	In-scope Sample	Realized Number of Responding Graduates		Response Rate (%)	
			Master	Share	Master	Share
Doctorate	654	649	245	236	37.8	36.4
Yukon	**143**	**136**	**62**	**57**	**45.6**	**41.9**
College diploma	107	101	52	47	51.5	46.5
Bachelor's degree	36	35	10	10	28.6	28.6
Northwest Territories	**181**	**176**	**52**	**42**	**29.5**	**23.9**
College diploma	180	175	52	42	29.7	24.0
Bachelor's degree	1	1	0	0	0	0
Nunavut	**125**	**124**	**38**	**33**	**30.6**	**26.6**
College diploma	125	124	37	32	29.8	25.8
Bachelor's degree	0	0	1	1	--	--
Canada	**59,467**	**58,467**	**28,715**	**27,155**	**49.1**	**46.4**
Trade/vocational	5,810	5,716	2,550	2,372	44.6	41.5
College diploma	17,659	17,030	8,712	8,111	51.2	47.6
Bachelor's degree	19,035	18,895	9,423	8,982	49.9	47.5
Master's degree	11,444	11,342	5,630	5,403	49.6	47.6
Doctorate	5,519	5,484	2,400	2,287	43.8	41.7

The global response rate of 49.1% (on the Master file) is a slight underestimate of the true response rate. In calculating this rate, we essentially split our Sample Size Sent to Collection into those that responded, those that were found to be out of scope, and those that did not respond. In reality, some of the non-responding units would also have been found to be out of scope, effectively increasing the response rate. We can estimate the magnitude of the difference:

Out of the 29,715 "resolved" units, we found that 1,000 were out of scope, and 28,715 were respondents; in other words, 3.37% of the resolved units were out of scope. So, it's reasonable to project that the same percentage of the 29,752 non-responding units, or another 1,001 units, would also have been out of scope. If we remove this number from the "In-scope Sample", we find that a more realistic estimate of the global (Master) response rate would be 28,715 / 57,466, or 50.0%.

For the PUMF, a subsample of 14,745 units was drawn from the 28,715 respondents.

9.0 Treatment of Non-response and Weighting

The National Graduates Survey – Class of 2009/2010 (NGS2013) is a probability survey. As is the case with any probability survey the sample is selected to represent a reference population - the graduate population - at a specific date within the context of the survey as accurately as possible. Each unit in the sample must therefore represent a certain number of units in the population. If the frame used was perfect (covering exactly the population of interest) and all selected units were traced, contacted and completed the survey, then the design weight assigned to each unit would represent accurately and exactly the number of graduates in the target population. In this situation, using this weight would yield unbiased estimates. However, this is not the case when surveys are faced with non-response and imperfect frames. Weight adjustments are traditionally used to compensate for these different issues. Response patterns have to be studied carefully to appropriately correct for non-response.

It was observed that non-response did not occur randomly or uniformly within the population, since as shown in Table 8.1, different response rates were obtained for different geographical regions or levels of certification. The use of appropriate techniques will correct non-response bias that may be introduced. They are summarized here, and described in a bit more detail in the following sections.

The chosen technique for the NGS2013 was based on response homogeneous groups (RHGs). RHGs are developed with the premise of identifying factors that influence the likelihood to respond, and then grouping together the sample units with similar profiles as defined by these factors. Then, the weights of the non-responding units in an RHG are redistributed among the responding units in the same RHG. The factors used to define these RHGs can come from data from the frame, or from the collection process itself.

Upon contacting the respondent, we sometimes found that the stratification data on the frame was incorrect; the field of study was particularly susceptible to error. These units, often referred to as stratum jumpers, will usually have an initial design weight different than other units in their (new) stratum. Furthermore, the RHGs for non-response adjustment were formed using more than just the stratification variables. As a result, the weights varied substantially, sometimes within one stratum. Verification and validation exercises led to some weights being modified to mitigate the impact of the most influential observations.

Final weights on the Master file were used as the basis for the PUMF weights. Respondents to the survey that were not selected into the PUMF were treated as non-respondents: their weights were redistributed among the selected units.

9.1 Design Weight

At the time of selection, an initial design weight was assigned to each graduate as the inverse of its probability of selection. Since the NGS2013 design is stratified with simple random sampling within strata, the probability of selection of the graduate i in stratum h is:

$$\pi_{ih}^{design} = \frac{n_h}{N_h}$$

where n_h and N_h denote respectively the sample and population size of stratum h.

Therefore, the design weight is:

$$w_{ih}^{design} = \frac{1}{\pi_{ih}^{design}}$$

9.2 Non-response adjustment

After the calculation of the design weight, a non-response adjustment was applied on the sample units. The sample was divided into two groups: resolved units and unresolved units. The group of resolved units contains the survey respondents and the out-of-scope units identified at collection (e.g. those whose graduation dates were not in 2009/2010). The group of unresolved units contains the rest of the sample, i.e., the non-respondents. For simplicity, we use the term non-response adjustment but in fact, it is an unresolved adjustment. For the purpose of this adjustment, response homogeneity groups (RHGs) were formed. RHGs are determined through a combination of logistic regressions to predict the probability of being a resolved unit and then using a clustering procedure based on the modelled probability of being a resolved unit. For building the logistic regression model, the explanatory variables considered included the stratification variables, demographic variables from the frame, and variables derived from the collection process. The variables retained for the logistic model were the province of study, certification level, field of study, citizenship, provincial mobility (whether their province of residence and province of study are the same), age group, and the number of tracing sources that were made available to interviewers.

For graduate i in RHG g the non-response adjustment is:

$$\pi_{ig}^{nonresp} = \frac{\sum_i w_{ih}^{design} I_{ig}}{\sum_i w_{ih}^{design} I_{ig} I_{ir}}$$

where I_{ig} equals 1 if graduate i is in RHG g ; equals 0 otherwise.

I_{ir} equals 1 if graduate i is resolved and in RHG g ; equals 0 otherwise.

So the non-response-adjusted weight becomes:

$$w_i^{adj} = w_{ih}^{design} \times \pi_{ig}^{nonresp}$$

9.3 Subsampling adjustment for the PUMF

The PUMF subsample was selected from the NGS2013 respondents, stratified according to their final weight on the Master File. The selection was done randomly within strata, and the non-selected units were simply treated as non-respondents. The adjustment to get the PUMF weight was therefore:

$$w_i^{pumf} = w_i^{master} \times \pi_{ih}^{pumf}$$

where $\pi_{ih}^{pumf} = \frac{\sum_i w_{ih}^{master} I_{ih}}{\sum_i w_{ih}^{master} I_{ih} I_{ip}}$, where I_{ih}=1 if unit i is in PUMF stratum h, and is 0 otherwise, and I_{ip}=1 if unit i was selected into the PUMF in stratum h, and is 0 otherwise.

9.4 Post-stratification and Validation

A few inaccuracies on the frame were discovered only during the collection process. For example, the initial frame included young adults taking short courses (e.g. hunting safety), as well as foreign students that had done their studies remotely. Such units shouldn't have been on the frame, so they were immediately coded as out of scope and removed from the collection process. The same criteria used to remove them from collection were used to remove them from the frame. The stratum jumpers also provided an opportunity to update the totals on the frame. The non-response-adjusted weights were therefore re-calibrated such that they would yield the updated frame totals.

$$w_i^{master} = w_i^{adj} \times \pi_{ik}^{post}$$

where π_{ik} is the adjustment factor for graduate *i* in post-stratification group *k*

PUMF weights were also post-stratified in a similar fashion. The post-stratification groups were defined by the combination of REGONID and CERTLEVP, the aggregated geographical regions and certification levels that are available on the PUMF.

9.5 Adjustment for Non-sharing Units

A Share File containing only the respondents who agreed to share their data was also created. Approximately 95% of the respondents agreed to share their data. After noting that the refusal to share was essentially random (no particular groups had a larger propensity to share), it was decided that the non-sharing adjustment would be applied within each stratum, as defined by the province of study, the level of certification and the field of study. The adjustment was treated much like a non-response adjustment: all of the sharing units in the stratum absorbed the weights of the non-sharing units. Since the reweighting was done within strata, no further post-stratification was needed, and validation exercises did not reveal any new needs for weight adjustments.

For graduate i in stratum *h*, the non-sharing adjustment is:

$$\pi_{ih}^{share} = \frac{\sum_i w_i^{master} I_{ih}}{\sum_i w_i^{master} I_{ih} I_{ishare}}$$

where I_{ih} equals 1 if graduate i is in stratum *h*; equals 0 otherwise.

I_{ishare} equals 1 if graduate i is a sharer and in stratum *h*; equals 0 otherwise.

The share weight consists of multiplying the master weight and the non-sharing adjustment.

For graduate i the share weight is:

$$w_i^{share} = w_i^{master} \times \pi_{ih}^{share}$$

The share weight is named WTPS on the NGS2013 Share File.

10.0 Data Quality

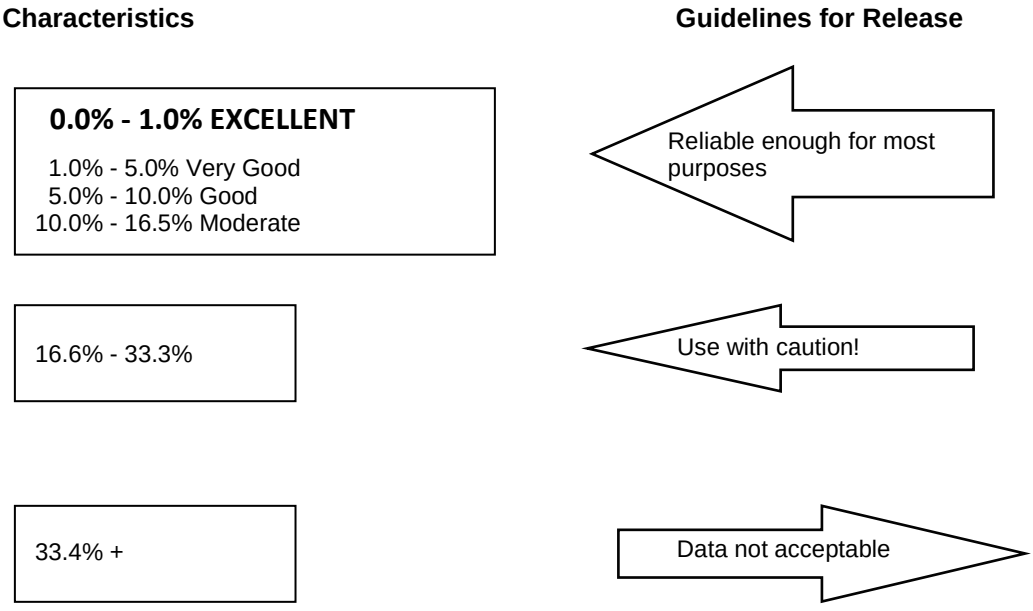
This chapter provides the user with information about the various factors affecting the quality of the survey data. There are two main types of errors: sampling errors and non-sampling errors. A sampling error is the difference between an estimate derived from a sample and the one that would have been obtained from a census that used the same procedures to collect data from every person in the population. All other types of errors such as frame coverage, response, processing and non-response are non-sampling errors. Many of these errors are difficult to identify and quantify. These are discussed in Section 10.2.

10.1 Sampling Errors

The estimates derived from the National Graduates Survey – Class of 2009/2010 (NGS2013) are based on a sample of graduates and not from a complete enumeration (census). This difference is the sampling error of the estimates.

The basis for measuring sampling error 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 measure, known as the coefficient of variation (CV) of an estimate, is obtained by expressing the standard error of the estimate as a percentage of the estimate. This measure allows for better quality comparisons between different types of estimates. The smaller the CV, the smaller the sampling variability, meaning smaller CVs are more desirable. The CV depends on the size of the sample on which the estimate is based, the population size and on the distribution of the sample, i.e. the sampling fraction of the units of the domains being estimated. The following diagram presents the characteristics of some CVs and the Statistics Canada guidelines for release.

Note that for the NGS2013, the error due to non-response has been incorporated into the sampling error. As described in Section 10.2 the use of the Generalized Estimation System (GES) takes into account the non-response variability into the estimates variability.



10.2 Non-sampling Errors

There are many sources of non-sampling errors that are not related to sampling, but may occur at almost any phase of a survey operation. Interviewers may misunderstand survey instructions, graduates may make a mistake in answering the questions, responses may be recorded in the questionnaire incorrectly or errors may be made in the processing or tabulating of the data. For the NGS2013, quality assurance measures were implemented at each phase of the data collection to monitor the quality of the data. These measures included precise interviewer training with respect to the survey procedures and questionnaire, observation of interviews to detect questionnaire design problems or misinterpretation of instructions and coding and edit quality checks to verify the processing logic. Chapter 7.0 outlines data processing procedures. Other kinds of non-sampling error are more easily quantifiable, especially non-response and population frame under-/over-coverage, the topics of the next two sections.

10.3 Non-response

Non-response, if not appropriately corrected, is a type of error that can lead to bias in the survey estimates. For the NGS2013, non-response significantly reduced the number of usable records. Biased estimates can occur when unusable units have significantly different characteristics from the usable ones. In Chapter 8.0, non-response rates were computed for basic domains to describe its extent. Extensive studies were completed on non-response to construct the proper adjustment weights for the NGS2013. Since the use of the final weights will yield the appropriate estimates of the population counts and ensure that non-respondents are incorporated and accounted for, it stresses the importance of using the final weights in any tabulations or analysis using the NGS2013 data. Any estimation done without the use of weights may produce biased or incorrect results.

Note that the census of graduates in some strata does not mean that no errors occurred and that the resulting variance will be zero in these strata. As mentioned in the previous section, the variance due to non-response is accounted for in the calculation of the final weight. Consequently, the resulting CVs reflect the global quality of the estimates even for units collected from a census.

10.4 Coverage

Coverage is an indication of how a survey frame covers the target population. There could be over-coverage if the survey frame contains units that should not have been included, such as deaths, duplicates, or incorrect date of graduation captured on the file. There could also be under-coverage, if the survey frame missed some units that should have been included.

For the NS2013, there was some under-coverage for graduates of colleges in some provinces. Data required to build the frame could not be obtained from a few institutions and therefore, graduates from those institutions were not included on the frame. Consequently, they could not be selected nor represented in any tabulation. No adjustment was made at the weighting stage to compensate for this under-coverage.

11.0 Guidelines for Tabulation Analysis and Release

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

11.1 Rounding Guidelines

In order that estimates for publication or other release derived from the National Graduates Survey – Class of 2009/2010 (NGS2013) microdata file correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

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

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

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

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

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

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

11.2 Sample Weighting Guidelines for Tabulation

The NGS2013 uses a stratified simple random sample design without replacement of graduates within strata. When producing simple estimates, including the production of ordinary statistical tables, users must use the final weight associated with the graduates concerned by the analysis. If final weights are not used, the estimates derived from the microdata file cannot be considered to be representative of the survey population and will not correspond to those produced by

Statistics Canada. The final weight assigned to a given responding graduate reflects the number of graduates in the NGS2013's population he/she represents.

For any analysis dealing with correlation analysis or any other statistics where a significance measure is required, it is recommended that an adjusted weight be used. This weight is obtained by multiplying the final weight by the sample size and dividing this total by the total estimated population. This produces a mean weight of 1 and a sum of weights equal to the sample size.

The benefit of this adjusted weight is that an overestimation of the significance (which is very sensitive to sample size) is avoided while maintaining the same distributions as those obtained when using the demographic weight. The disadvantage is that the numerator is not weighted up to the target population and the coefficient of variance is no longer useful as a measure of data quality.

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.

11.3 Definitions of Types of Estimates: Categorical and Quantitative

The NGS2013 file has been set up so that the graduate is the unit of analysis. The final weight that can be found on each record is called WTPM (WTPS for the share file) in the codebook.

The unit of analysis for the NGS2013 PUMF is also the graduate, and the final PUMF weight is called WTPP in the codebook.

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 or the proportion of self-employed graduates working at a job last week is an example 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 a business?
R: Yes / No

Q: At your (main) job last week, were you a paid worker or self-employed?
R: Paid worker / Self-employed / Unpaid family worker

Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X}/\hat{Y}$ where $\hat{X}$ is an estimate of surveyed population quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of hours worked per week at a job. The numerator is an estimate of the total number of hours worked per week and its denominator is the number of graduates working.

Examples of Quantitative Questions:

Q: How many (paid) hours a week do you usually work at this job?
R: |_|_|_| hours

Q: How much do you now owe for all your government-sponsored student loans?
R: |_|_|_|_|_|_| dollars

11.3.1 Tabulation of Categorical Estimates

Estimates of the number of graduates with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form $\hat{X}/\hat{Y}$ are obtained by:

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

11.3.2 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 job in the week before they were surveyed multiply the value reported in question LF_Q79 (hours worked per week) by the final weight for the record, then sum this value over all records with LFSTAT = 1 (employed) and LF_Q79 < 996.

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of hours worked by graduates in their main job in the week before they were surveyed,

a) estimate the total number of hours ($\hat{X}$) as described above,
b) estimate the number of graduates ($\hat{Y}$) in this category by summing the final weights of all records with LFSTAT = 1 and LF_Q79 < 996, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

11.4 Guidelines for Statistical Analysis

The NGS2013 is based upon a sample design with stratification and different probabilities of selection, depending on the stratum and non-uniform non-response patterns. Using data from such surveys presents problems to analysts because the survey design items mentioned above affect the estimation and variance calculation procedures that should be used. For all types of analysis, final weights are strongly suggested.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures may differ from that which is appropriate in a sample survey framework, with the result that, while in many cases the estimates produced by the packages are correct, the variance estimates that are calculated are poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables and for common domains) can be derived using the accompanying Approximate Sampling Variability Tables (see Chapter 12.0). Also, approximate release cut-offs have been calculated and are presented in Section 11.6.

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.

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

11.5 Release Guidelines

Before releasing and/or publishing any estimate from NGS2013, users should first determine the quality level of the estimate. The quality levels are *acceptable*, *marginal* and *unacceptable*. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 10.0.

First, the number of graduates (unweighted) who contribute to the calculation of the estimate should be determined. If this number is less than 5, the weighted estimate should be considered of unacceptable quality and more importantly too small for disclosure. Users are invited to read the document Statistics Canada Quality Guidelines available on Statistics Canada web site.

Once this criterion is met, users must determine the coefficient of variation of the estimate and follow the guidelines below. All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users. These quality level guidelines should be applied to weighted rounded estimates.

Quality Level Guidelines

Quality Level of Estimate	Guidelines
1) Acceptable	Estimates have a sample size of five graduates or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of five graduates or more, and high coefficients of variation in the range of 16.6% to 33.3%.

Quality Level of Estimate	Guidelines
	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 five graduates, 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: "Please be warned that these estimates [flagged with the letter U] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

11.6 Release Cut-offs

The tables in Appendix I provide an indication of the precision of population estimates as they show the release cut-offs associated with a CV of 16.5% and a CV of 33.3% (correspond to quality levels presented in the previous section). These cut-offs are derived from the ***Approximate Sampling Variability Tables*** discussed in Chapter 12.0. For example, Appendix I shows that the quality of a weighted estimate of 2,000 graduates in Ontario possessing a given characteristic is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 12.1 and Example 4 in Section 12.1.1 explain the correct procedure to be used for ratios.

12.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation (CV) that would be applicable to a wide variety of categorical estimates produced from the NGS2013 microdata file, and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced (see Appendix J). These tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation are derived using the variance formula for simple random sampling, and incorporating a factor which reflects the sample design and the adjustment for non-response. This factor, known as the design effect, was determined by first calculating design effects for a wide range of

characteristics, and then choosing from among these a conservative value (usually the 75th percentile) to be used in the CV tables, which would then apply to the entire set of characteristics.

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

Note: As CVs provided in the Approximate Sampling Variability Tables are approximate, it is not recommended to use these tables if the estimate is based on a weighted number of observations less than 30. This is the reason why no CV is given in the tables when that number is less than 30. Also, the user must follow the other release guidelines presented in Section 11.5.

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

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

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

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

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

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage, and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of males among all foreign student graduates is more reliable than the estimated number of male foreign student graduates. (Note that in the tables the coefficients of variation decline in value reading from left to right).

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

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

Rule 3: Estimates of Differences Between Aggregates or Percentages

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

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

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

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of male graduates and the numerator is the number of male foreign student graduates.

In cases where the numerator is not a subset of the denominator, for example, the ratio of the number of male graduates in Education as compared to the number of female graduates in Education, the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

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

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

12.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples based on the NGS2013 PUMF are included to assist users in applying the above rules.

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

Suppose that a user estimates that 20,909 graduates were in a program that included components taken outside of Canada. How does the user determine the coefficient of variation of this estimate?

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

2) The estimated aggregate (20,909) 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 20,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, in this case 6.0%.

4) So the approximate coefficient of variation of the estimate is 6.0%. The finding that there are 20,909 graduates (to be rounded according to the rounding guidelines in Section 11.1) who were in a program that included components taken outside of Canada is publishable with no qualifications.

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

Suppose that the user estimates that 9,284 / 20,909 = 44.4% of graduates who took program components outside of Canada are married or in common-law relationships. How does the user determine the coefficient of variation of this estimate?

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

2) The numerator, 9,284, 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 9,000. Similarly, the percentage estimate does not appear in any of the column headings, so it is necessary to use the percentage closest to it, 40%.

3) The figure at the intersection of the row and column, 7.3%, is the coefficient of variation to be used.

4) So the approximate coefficient of variation of the estimate is 7.3%. The finding that 44.4% of graduates who took program components outside of Canada are married or in common-law relationships can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 3,157 / 7,590 = 41.6% of male graduates who took program components outside of Canada are married or in common-law relationships, while 6,126 / 13,318 = 46.0% of female graduates who took program components outside of Canada are married or common-law. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for men as 12.7%, and the CV of the estimate for women as 8.2%.

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

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

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

That is, the standard error of the difference $\hat{d} = 0.460 - 0.416 = 0.044$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.460)(0.082)]^2 + [(0.416)(0.127)]^2} \\ &= \sqrt{(0.0014227) + (0.0027912)} \\ &= 0.065\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} =$ 0.065 / 0.044 = 1.48%

4) So the approximate coefficient of variation of the difference between the estimates is 148%. The quality of the estimated difference is considered unacceptable, and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter U (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

Example 4: Estimates of Ratios

Suppose that the user estimates that 42,325 males supervised other employees at their main job last week, while 45,594 females supervised other employees at their main job last week. The user is interested in comparing the estimate of men versus women in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) This is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of male graduates who supervised other employees at their main job last week. The denominator of the estimate ($\hat{X}_2$) is the number of female graduates who supervised other employees at their main job last week.

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

3) The numerator of this ratio estimate is 42,325. The figure closest to it is 42,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely 4.1%.

4) The denominator of this ratio estimate is 45,594. The figure closest to it is 50,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, 3.7%

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

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

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

That is:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.041)^2 + (0.037)^2} \\ &= \sqrt{0.001681 + 0.001369} \\ &= 0.055\end{aligned}$$

6) The obtained ratio of male versus female graduates who supervised other employees at their main job last week is 42,325 / 45,594, which is 0.93 (to be rounded according to the rounding guidelines in Section 11.1). The coefficient of variation of this estimate is 5.5%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of male to female graduates who supervised other employees at their main job last week, is 0.93 at the Bachelor level (CERTLEVP=2) and 0.87 at the Master/Doctorate level (CERTLEVP=3). The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the Bachelor ratio ($\hat{R}_1$) and the Master/Doctorate ratio ($\hat{R}_2$) as in Example 4. The approximate CV is 8.1% for the Bachelor ratio, and 14.4% for the Master/Doctorate ratio.

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

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

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 0.93 – 0.87 = 0.06 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.93)(0.081)]^2 + [(0.87)(0.144)]^2} \\ &= \sqrt{(0.0056746) + (0.015695)} \\ &= 0.146\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.146 / 0.06 = 2.43

4) So the approximate coefficient of variation of the difference between the estimates is 243%. The quality of the estimated difference is considered unacceptable, and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter U (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

12.2 How to Use the Coefficient of Variation Tables to Obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example, a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the difference would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X} - k, \hat{X} + k\right)$ where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{x}}$):

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ is the determined coefficient of variation of $\hat{X}$, and

$t = 1$ if a 68% confidence interval is desired;
$t = 1.6$ if a 90% confidence interval is desired;
$t = 2$ if a 95% confidence interval is desired;
$t = 2.6$ if a 99% confidence interval is desired.

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

12.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of graduates who are married or in common-law relationships among those who took program components outside of Canada (from Example 2, Section 12.1.1) would be calculated as follows:

$\hat{X}$ = 44.4% (or expressed as a proportion 0.444)

t = 2

$\alpha_{\hat{x}}$ = 7.3% (0.073 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.444 – (2)(0.444)(0.073), 0.444 + (2)(0.444)(0.073)}

$CI_{\hat{x}}$ = {0.444 – 0.065, 0.444 + 0.065}

$CI_{\hat{x}}$ = {0.379, 0.509}

With 95% confidence, it can be said that between 37.9% and 50.9% of graduates who took program components outside of Canada are married or in common-law relationships.

12.3 How to Use the Coefficient of Variation Tables to Do a T-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. In other words, the difference between the estimates is significant.

12.3.1 Example of Using the Coefficient of Variation Tables to Do a T-test

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of male and female graduates who are married or in common-law relationships among those who took program components outside of Canada. From Example 3, Section 12.1.1, the standard error of the difference between these two estimates was found to be 0.065. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.460 - 0.416}{0.065} = \frac{0.044}{0.065} = 0.677$$

Since t = 0.677 is between -2 and 2, no conclusion about the difference between the characteristics is justified at the 5% level of significance.

12.4 Coefficients of Variation for Quantitative Estimates

Special tables would have to be produced to determine the sampling error of quantitative estimates. Since most of the variables for the NGS2013 PUMF are categorical in nature, this has not been done.

As a general rule, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the estimated number of persons contributing to it.

12.5 Coefficient of Variation Tables

Approximate Sampling Variability Tables are available in Appendix J.

13.0 Questionnaire and Code Sheets

Please refer to the files listed below for the National Graduates Survey **–** Class of 2009/2010 (NGS2013).

Questionnaire:
NGS_2013_Questionnaire.doc
NGS_2013_Questionnaire.pdf

Code Sheets:

Classification of Instructional Programs (CIP 2000)
NGS_2013_Appendix_A1_CIP_2000.doc
NGS_2013_Appendix_A1_CIP_2000.pdf

Classification of Instructional Programs (CIP 2011)
NGS_2013_Appendix_A2_CIP_2011.doc
NGS_2013_Appendix_A2_CIP_2011.pdf

North American Industry Classification System (NAICS) 2012
NGS_2013_Appendix_B_NAICS_2012.doc
NGS_2013_Appendix_B_NAICS_2012.pdf

National Occupational Classification for Statistics (NOC) 2011
NGS_2013_Appendix_C_NOC_2011.doc
NGS_2013_Appendix_C_NOC_2011.pdf

Citizenship Codes
NGS_2013_Appendix_D_Citizenship_codes.doc
NGS_2013_Appendix_D_Citizenship_codes.pdf

Country and Immigration Codes
NGS_2013_Appendix_E_Country_Immigration_Codes.doc
NGS_2013_Appendix_E_Country_Immigration_Codes.pdf

Language Codes
NGS_2013_Appendix_F_Language_Codes.doc
NGS_2013_Appendix_F_Language_Codes.pdf

State Codes
NGS_2013_Appendix_G_State_Codes.doc
NGS_2013_Appendix_G_State_Codes.pdf

Derived variables
NGS_2013_Appendix_H_Derived_Variables.doc
NGS_2013_Appendix_H_Derived_Variables.pdf

Release cut-off tables for estimates by various domains
NGS_2013_Appendix_I_Release_Cut_off_Tables.doc
NGS_2013_Appendix_I_Release_Cut_off_Tables.pdf

Approximate Sampling Variability Tables by various domains
NGS_2013_Appendix_J_Approximate_Sampling_Variability_Tables.doc
NGS_2013_Appendix_J_Approximate_Sampling_Variability_Tables.pdf

14.0 Record Layout with Univariate Frequencies

See NGS2013_Master_CdBkE.doc or NGS2013_Master_CdBkE.pdf for the record layout with univariate counts for the master file.

See NGS2013_Share_CdBkE.doc or NGS2013_Share_CdBkE.pdf for the record layout with univariate counts for the share file.

See NGS2013_PUMF_CdBkE.doc or NGS2013_PUMF_CdBkE.pdf for the record layout with univariate counts for the master file.