Microdata User Guide

TRAVEL ACTIVITIES AND MOTIVATION SURVEY

2006

Table of Contents

1.0   Introduction

The Travel Activities and Motivation Survey (TAMS) was conducted by Statistics Canada in 2006 with the cooperation and support of eight provincial and territorial ministries and agencies responsible for tourism as well as the Canadian Tourism Commission, Parks Canada, Canadian Heritage and the Atlantic Tourism Partnership. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

Statistics Canada

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

Ontario Ministry of Tourism

Alex Athanassakos
Ontario Ministry of Tourism
700 Bay Street, 15th Floor
Toronto, Ontario M7A 2E1
Telephone: (416) 314-7317
Fax: (416) 314-7341
E-mail: Alex.Athanassakos@ontario.ca

2.0   Background

The 2006 Travel Activities and Motivation Survey (TAMS) was conducted between January and June 2006 to collect information on Canadians' travel habits, participation in recreational activities and motivators to travel.

This two phase survey was comprised of an initial computer-assisted telephone interview (CATI) to identify travellers and non-travellers and a follow-up mail-out/mail-back paper questionnaire to travellers.

The questionnaire for the 2006 TAMS was modified from the previous TAMS, conducted in 1999.

Data from the TAMS are used by provincial ministries of tourism as well as federal government agencies and departments. Other users include the media, business, consultants, universities and other researchers interested in Canadian travellers.

3.0   Objectives

The survey's overall objective is to collect information on Canadians' travel activities and motivation to travel.

Other objectives of the Travel Activities and Motivation Survey are:
- to collect information on out-of-town trips of one or more nights taken in the past two years in Canada, the USA and other countries;
- to collect information on the types of recreational and entertainment activities undertaken while travelling;
- to profile travel experiences and motivators by socio-demographic factors;
- to identify types of travel activities that motivate travel;
- to examine the relationship between travelling, participation in activities and destinations;
- to collect information on vacation planning;
- to better understand reasons for not travelling.

4.0   Survey Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaires used.

Household member
A household member is any person who, at the time the roster is completed:
- considers or is reported to consider the dwelling as their usual place of residence; or
- is staying in the dwelling and has no other usual place of residence elsewhere

This includes:
- a spouse, partner or child temporarily away from home due to work or school but who considers this as his/her usual place of residence and who has resided in this dwelling for a minimum of 30 days in the past 12 months;
- children in joint physical custody;
- a person temporarily residing in an institution who has been absent from his/her dwelling for less than six months;
- a person applying for refugee status;
- a student attending school in Canada on a student visa; and
- a person in Canada on a work permit.

Selected respondent
The selected respondent is the household member, 18 years or older, who has been randomly chosen during the telephone interview to complete the survey.

Traveller
For the Travel Activities and Motivation Survey, a traveller is defined as someone who has taken an out-of-town trip of one or more nights away from home during the past two years.

Main activity
Main activity is the activity where the respondent spends most of his/her time.

Paid vacation days
Paid vacation days are the number of paid days off from work that a person earns each year from a paid job. The person may or may not have used all of their vacation days in the year.

Working at a job or business
Working at a job or business means that a person is either a paid employee, self-employed in his/her own business, trade or profession, or an unpaid employee in a family business or farm. This includes any activity carried out by the respondent for pay or profit including part-time work, and "payment in kind" (payment in goods or services rather than money). Work around the house or volunteer work, such as for a church, is not counted as working at a job or business.

Self-employed
A person is self-employed when he/she earns income directly from his/her own business, trade or profession, rather than being paid a specified wage or salary by an employer.

5.0   Survey Methodology

The telephone survey was carried out from January to April 2006 using a Random Digit Dialling (RDD) telephone sampling method. The follow-up mail-out/mail-back survey, conducted between January and June 2006, used the addresses obtained during the telephone survey to contact travellers.

5.1   Population Coverage

The target population for the telephone survey was all persons 18 years of age and older in each of the 10 Canadian provinces, excluding full-time residents of institutions. Because the survey was conducted using a sample of telephone numbers, households (and thus persons living in households) that do not have a telephone land line were excluded from the sample population. This means that people without telephones and people with cell phones only, were excluded. People without land lines account for less than 6% of the target population. However, the survey estimates have been adjusted through weighting to represent persons without land lines.

The target population for the mail survey was all persons 18 years of age and older in each of the 10 Canadian provinces, excluding full-time residents of institutions and non-travellers, who had taken an out-of-town trip of one or more nights during the past two years. Travellers were identified through a screening question during the telephone interview.

5.2   Sample Stratification

The sample was stratified at the census metropolitan area (CMA) level, as follows:
- in the Atlantic Provinces: Halifax, Saint John, St. John's, Other Atlantic;
- in Quebec: Montreal, Quebec City, Gatineau, Other Quebec;
- in Ontario: Toronto, Ottawa, Hamilton, London, Kitchener, St. Catharines-Niagara, Windsor, Oshawa, Greater Sudbury, Kingston, Thunder Bay, Other Ontario;
- in Manitoba: Winnipeg, Other Manitoba;
- in Saskatchewan: Saskatoon, Regina, Other Saskatchewan;
- in Alberta: Edmonton, Calgary, Other Alberta; and
- in British Columbia: Vancouver, Victoria, Abbotsford, Other British Columbia.

5.3   Random Digit Dialling Sample Selection

The Travel Activities and Motivation Survey (TAMS) sample was selected using Random Digit Dialling, a technique whereby telephone numbers are generated randomly by computer. The method uses the concept of banks. Every Canadian telephone number is made up of an area code, a prefix and four digits. The area code, prefix and the next two digits define the hundreds bank. For example, the telephone number 613-555-1234 belongs to area code 613, prefix 555 and bank 61355512.

The RDD frame consists of working banks compiled from telephone company administration files. A working bank, for the purposes of social surveys, is defined as a bank which contains at least one working residential telephone number. Thus, all banks with only unassigned, non-working, or business telephone numbers are excluded from the RDD frame. In Canada, there are about 269,000 working banks (over 26 million possible numbers). Each of the banks are assigned to a province and within a province, to a CMA or to the non-CMA portion of the province. The assignment is based on the bank's area code and prefix (ACP).

The TAMS RDD sample was selected in several steps: The first step involved selecting a large simple random sample with replacement (SRSWR) of banks, within each CMA and non-CMA

listed in Section 5.2. For each selected bank, the last two digits needed to complete the telephone number was chosen at random from among the 100 possibilities: 00 to 99. The generated telephone numbers were then stratified into three strata based on their status: residential, business or unknown. The final step involved selecting a simple random sample without replacement (SRSWOR) from the residential stratum, and another from the "unknown" stratum within each CMA and non-CMA.

5.4 Sample Design

The RDD sample is a stratified simple random sample of telephone numbers selected with replacement. The sample was stratified by CMA (see Section 5.2) by telephone status (residential/unknown). A screening activity aimed at removing not in service numbers was performed for telephone numbers of "unknown" status prior to sending the sample to the computer-assisted telephone interviewing (CATI) unit.

Each telephone number in the CATI sample was dialled to determine whether or not it reached a household. If the telephone number was found to reach a household, the person answering the phone was asked to list all members in the household and to provide the age and sex of each member. One member 18 years of age or older was randomly selected to participate in the survey. The ultimate sampling unit is the selected person.

All respondents identified as travellers during the telephone interview were asked to complete a mail-out/mail-back paper questionnaire. The sample design for the mail-out/mail-back survey is a two phase design: the first phase is the telephone interview, and the second phase is the paper questionnaire for travellers only.

5.5   Sample Size

The initial RDD sample consisted of 132,065 telephone numbers nationally, allocated to the strata as given in the table below. To increase the proportion of productive numbers in the initial sample, telephone numbers with a residential status were over-sampled, compared to telephone numbers with an "unknown" status. The tables in Section 8.1 provide information on the number of respondents and the response rates.

Initial Sample Size by Stratum

Census Metropolitan Areas	Residential	Unknown	Total
St. John's	663	455	1,118
Halifax	2,169	1,357	3,527
Saint John	642	568	1,210
Other Atlantic	3,090	1,778	4,868
Quebec City	3,556	938	4,493
Montreal	6,613	2,174	8,787
Gatineau	3,475	1,037	4,512
Other Quebec	5,491	2,181	7,672
Ottawa	3,202	1,531	4,734
Kingston	644	170	813
Oshawa	732	176	908
Toronto	12,426	6,262	18,688
Hamilton	2,994	1,146	4,141
St.Catharines-Niagara	924	401	1,325
Kitchener	2,279	840	3,119
London	2,705	1,089	3,794
Windsor	837	297	1,134
Greater Sudbury	817	420	1,237
Thunder Bay	539	293	831
Other Ontario	5,939	2,871	8,810
Winnipeg	2,699	1,126	3,826
Other Manitoba	2,664	1,278	3,942
Regina	2,229	899	3,128
Saskatoon	2,178	689	2,867
Other Saskatchewan	2,476	1,200	3,676
Calgary	3,163	863	4,026
Edmonton	3,119	766	3,885
Other Alberta	3,292	1,186	4,479
Abbotsford	845	193	1,037
Vancouver	6,372	1,748	8,120
Victoria	2,683	480	3,163
Other British Columbia	3,442	754	4,196
Canada	**94,898**	**37,168**	**132,065**

6.0   Data Collection

Data collection for the Travel Activities and Motivation Survey (TAMS) was carried out from January to June 2006. It consisted of two phases: a telephone survey to identify travellers and non-travellers and a mail-out/mail-back survey completed by travellers.

6.1   Interviewing

The telephone survey was conducted using computer-assisted interviewing (CAI). Data collection for the telephone survey took place from January to mid-April 2006. Interviewing was administered through the Statistics Canada Regional offices in Halifax, Sherbrooke, Sturgeon Falls, Winnipeg and Edmonton.

Statistics Canada staff working on the TAMS, including project supervisors, senior interviewers and interviewers, participated in a training video conference designed to familiarize them with the objectives and concepts of the survey, the CAI questionnaire and procedures specific to the TAMS. An interviewer's manual was provided to support interviewers during the data collection.

Participation in the survey was voluntary. Proxy responses on behalf of the selected respondent were not permitted.

After all attempts had been made to interview a selected respondent, the case was assigned a final status code and returned to head office.

Data collection for the mail-out/mail-back survey took place from mid-January to mid-June 2006. This survey was administered through the Statistics Canada Operations and Integration Division. A telephone follow-up with non-responding travellers was conducted by staff of this division during the collection period.

6.2   Supervision and Quality Control

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

6.3   Non-response

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

7.0   Data Processing

The main output of the Travel Activities and Motivation Survey (TAMS) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this 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 encrypted to ensure confidentiality and are transmitted over a secure line to Ottawa for processing.

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

All of the mail-out/mail-back questionnaires were returned to the Operations and Integration Division at head office for data capture by a computerized Optical Character Recognition (OCR) system. Completed records were transferred electronically to Special Surveys Division for processing.

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 respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

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

7.3   Imputation

Imputation is the process that supplies valid values for variables that have invalid or missing data. Imputation was not appropriate for most items on either the telephone or the mail-out/mail-back questionnaire. Not stated codes were assigned to items with missing data. Records judged to have insufficient data were removed from processing.

The one variable that was subject to imputation is the census metropolitan area (CMA). Initial CMA codes were assigned during sample selection using the first six digits of the telephone number (the area code plus the prefix of the number, known as the ACP). Better quality CMA codes were derived from collected postal codes using the Postal Code Conversion File (PCCF). A validation process was implemented to verify the postal codes and derived CMA codes. Donor

imputation was performed for records with CMAs deemed invalid (1% of records) or missing postal codes (5% of records).

7.4   Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. These include age groups, education, country of birth, household composition and number of vacation days.

7.5   Weighting

The principle behind estimation in a probability sample such as the TAMS 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 travelling during the past two years 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 11.0.

7.6   Suppression of Confidential Information

It should be noted that the "Public Use" Microdata Files (PUMF) may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of file variables, grouping values into wider categories, and coding specific values into the "not stated" category. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

The survey master file includes the respondent's precise age, while the PUMF contains age groups only.

For certain variables that are susceptible to identifying individuals, the PUMF may have been treated with local suppression, that is, some of the values in the master file may have been coded as "not stated" on the PUMF.

8.0   Data Quality

8.1   Response Rates

The table below presents province level counts for the telephone survey.

Counts for the Telephone Component

Region	Total	Sent to the Field	Resolved in the Field	Estimated In-scope	Household Response	Person Response
Atlantic Provinces	10,722	8,910	8,200	6,291	4,631	4,170
Quebec	25,465	23,020	21,505	19,157	12,264	11,113
Ontario	49,534	43,910	40,698	33,989	20,085	17,961
Manitoba	7,767	6,721	6,062	5,382	3,498	3,189
Saskatchewan	9,671	8,277	7,464	6,948	4,620	4,270
Alberta	12,389	11,750	10,630	9,825	6,036	5,505
British Columbia	16,517	15,927	14,816	13,458	7,734	6,942
Canada	**132,065**	**118,515**	**109,375**	**95,051**	**58,868**	**53,150**

The columns in the table above are defined as follows:

Total
The total number of telephone numbers selected in the initial Random Digit Dialling (RDD) sample.

Sent to the Field
The number of telephone numbers sent to the field for collection. The difference between the "Total" column and the "Sent to the Field" column is the number of telephone numbers that were removed from the sample by a screening activity aimed at identifying not in service numbers prior to collection.

Resolved in the Field
The number of telephone numbers that were confirmed during collection as either in-scope (residential) or out-of-scope (e.g. business or non-working number).

Estimated In-Scope

$$= \sum_{\text{Stratum}} \left(\frac{\text{Number of in - scope}}{\text{Number of resolved in the field}} \text{ x Number sent to the field} \right)$$

In other words, the proportion of resolved telephone numbers confirmed in-scope was calculated within each stratum, and the same proportion was applied to the unresolved numbers.

Household Response
The number of cases with a complete household roster listing all the members of the household.

Person Response
The number of cases where the household member selected to participate in the survey provided sufficient useable data in the telephone interview to be considered a respondent.

The table below provides response rates based on the counts above.

Rates for the Telephone Component

Region	Overall Resolved Rate (%)	Field Resolved Rate (%)	Overall Hit Rate (%)	Field Hit Rate (%)	Household Response Rate (%)	Person Response Rate (%)	Overall Response Rate (%)
Atlantic Provinces	93.4	92.0	58.7	70.6	73.6	90.0	66.3
Quebec	94.1	93.4	75.2	83.2	64.0	90.6	58.0
Ontario	93.5	92.7	68.6	77.4	59.1	89.4	52.8
Manitoba	91.5	90.2	69.3	80.1	65.0	91.2	59.2
Saskatchewan	91.6	90.2	71.8	83.9	66.5	92.4	61.5
Alberta	91.0	90.5	79.3	83.6	61.4	91.2	56.0
British Columbia	93.3	93.0	81.5	84.5	57.5	89.8	51.6
Canada	**93.1**	**92.3**	**72.0**	**80.2**	**61.9**	**90.3**	**55.9**

The rates in the table above were calculated as follows:

$$\text{Overall Resolved Rate} = \frac{\text{Screened out numbers} + \text{Resolved in the field}}{\text{Total sample size}}$$

$$\text{Field Resolved Rate} = \frac{\text{Resolved in the field}}{\text{Sent to the field}}$$

$$\text{Overall Hit Rate} = \frac{\text{Estimated number of in-scope telephone numbers}}{\text{Total sample size}}$$

$$\text{Field Hit Rate} = \frac{\text{Estimated number of in-scope telephone numbers}}{\text{Sent to the field}}$$

$$\text{Household Response Rate} = \frac{\text{Household response}}{\text{Estimated number of in-scope telephone numbers}}$$

$$\text{Person Response Rate} = \frac{\text{Person response}}{\text{Household response}}$$

$$\text{Overall Response Rate} = \frac{\text{Person response}}{\text{Estimated number of in-scope telephone numbers}}$$

The next table provides counts and rates for the mail-out / mail-back paper questionnaire.

Counts and Rates for the Mail Component

Region	Telephone Travellers	Paper Sent	Percent Sent (%)	Paper Response	Response Rate (%)	Overall Response Paper (%)
Atlantic Provinces	3,495	3,361	96.2	1,806	53.7	51.7
Quebec	9,294	8,850	95.2	4,975	56.2	53.5
Ontario	15,569	15,253	98.0	8,153	53.5	52.4
Manitoba	2,683	2,599	96.9	1,561	60.1	58.2
Saskatchewan	3,786	3,598	95.0	2,064	57.4	54.5
Alberta	5,060	4,737	93.6	2,663	56.2	52.6
British Columbia	6,256	5,881	94.0	3,470	59.0	55.5
Canada	**46,143**	**44,279**	**96.0**	**24,692**	**55.8**	**53.5**

The columns in the table above were defined as follows:

Telephone Travellers
The number of travellers in the telephone survey.

Paper Sent
The number of paper questionnaires that were mailed out.

$$\text{Percent Sent} = \frac{\text{Paper sent}}{\text{Telephone traveller}}$$

Paper Response
The number of paper questionnaires received with sufficient useable data.

$$\text{Response Rate} = \frac{\text{Paper response}}{\text{Paper sent}}$$

$$\text{Overall Response Rate} = \frac{\text{Paper response}}{\text{Telephone traveller}}$$

The table below provides counts by census metropolitan areas (CMA).

Counts by Census Metropolitan Areas

Census Metropolitan Areas	Telephone Response	Telephone Travellers	Non-Travellers	Mail Response
Halifax	1,408	1,231	177	648
Other Atlantic	2,762	2,264	498	1,158
Quebec City	2,072	1,786	286	990
Montreal	3,642	3,103	539	1,565
Gatineau	2,063	1,756	307	979
Other Quebec	3,336	2,649	687	1,441
Ottawa	2,015	1,807	208	1,031
Kingston	437	387	50	222
Oshawa	399	342	57	180
Toronto	5,902	5,078	824	2,410
Hamilton	1,553	1,330	223	728
St.Catharines-Niagara	507	418	89	222
Kitchener	1,271	1,103	168	595
London	1,460	1,295	165	712
Windsor	431	372	59	181
Greater Sudbury	496	438	58	230
Thunder Bay	313	265	48	156
Other Ontario	3,177	2,734	443	1,486
Winnipeg	1,709	1,421	288	849
Other Manitoba	1,480	1,262	218	712
Regina	1,373	1,223	150	681
Saskatoon	1,386	1,227	159	694
Other Saskatchewan	1,511	1,336	175	689
Calgary	1,757	1,625	132	849
Edmonton	1,863	1,695	168	896
Other Alberta	1,885	1,740	145	918
Vancouver	2,996	2,671	325	1,384
Victoria	1,543	1,393	150	869
Other British Columbia	2,403	2,192	211	1,217
Canada	**53,150**	**46,143**	**7,007**	**24,692**

The TAMS public use microdata file (PUMF) contains the 7,007 non-travellers who responded to the telephone survey plus the 24,692 travellers who responded to the mail survey, for a total of 31,699 records. After receiving the paper questionnaires 609 respondents who were initially classified as travellers in the telephone interview had their status changed to non-traveller.

8.2   Survey Errors

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

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering

questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of <u>non-sampling errors.</u>

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

8.2.1   Frame Coverage

As mentioned in Section 5.1 (Population Coverage), less than 6% of households in Canada do not have telephone land lines. Individuals living in these households may have unique characteristics which will not be reflected in the survey estimates. Users should be cautious when analyzing subgroups of the population which have characteristics that may be correlated with non-telephone ownership or cell phone-only ownership.

8.2.2   Data Collection

Interviewer training consisted of reading the Travel Activities and Motivation Survey (TAMS) Supervisor's Manual, Procedures Manual and Interviewer's Manual, practicing with the TAMS training cases on the computer, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. The collection period for the TAMS telephone portion ran from January to April 2006.

The telephone survey and mail-out/mail-back questionnaire both asked a question regarding travel destinations in the past two years (question TS_Q02 for the telephone survey, and question A01 column A for the paper questionnaire). There were inconsistencies in the response category wording (i.e. *Europe (including UK and Russia)* on the paper questionnaire but only *Europe including the UK* on the telephone questionnaire) as well as missing categories (i.e. *Other countries* is an option on the telephone questionnaire but not the paper questionnaire).

There were also inconsistencies in respondents' data between the two sources of data: responses were compared for respondents who reported data for both TS_Q02 in the telephone survey and question A01 column A in the paper questionnaire. Unweighted counts are presented in the table below. For example, the number of respondents who reported that they had travelled within their own province for both questions is 17,630. Of those who reported that they had travelled within their own province in the telephone survey, 10.6% (2,083 / (17,630 + 2,083)) reported that they had not travelled within their own province in the paper questionnaire. Of those who reported that they had travelled within their own province in the paper questionnaire, 12.1% (2,438 / (17,630 + 2,438)) reported that they had not travelled within their own province in the telephone interview.

Question TS_Q02 (Telephone Interview) versus Question A01 column A (Paper Questionnaire)

	TS_Q02 = Yes & A01 = Yes	TS_Q02 = Yes & A01 = No		TS_Q02 = No & A01 = Yes		TS_Q02 = No & A01 = No
		Count	%	Count	%	
Own Province	17,630	2,083	10.6	2,438	12.1	2,535
Other Province	12,751	820	6.0	2,994	19.0	8,121
United States	10,333	716	6.5	2,063	16.6	11,574
Mexico	2,197	301	12.0	732	25.0	21,456
South/Central America	731	476	39.4	337	31.6	23,142
Caribbean	2,555	416	14.0	953	27.2	20,762
Europe	3,258	461	12.4	485	13.0	20,482
Asia	853	156	15.5	290	25.4	23,387

The inconsistencies between what was reported in the telephone interview and the paper questionnaire show evidence of non-sampling error. As well, for all destinations except South/Central America, the counts are higher for the paper questionnaire than for the telephone interview, which may suggest that respondents remember more destinations when filling out a questionnaire than when interviewed over the phone.

Another data quality issue related to data collection is that some variables were collected in the telephone interview for non-travellers, and in the paper questionnaire for travellers. Estimates for such variables will be based on data from two different modes of collection. Data users should also be aware that differences in the mode of collection are likely to have a non-negligible impact on analyses comparing travellers with non-travellers for these variables.

8.2.3   Data Processing

Data processing for the telephone survey was relatively straightforward since the data was captured using a computer-assisted telephone interview (CATI) application, in which edits and flows had been programmed to improve the consistency of the captured data. Data processing was much more complex for the paper questionnaire.

One of the tasks required for processing the paper questionnaire data was to create rules that determine whether to interpret blanks in the questionnaire as "No" or "Not stated". The section that caused the greatest difficulty was Section A, questions A03 to A17 since they do not have "None of the above" categories. The following rule was used: If there is a "Yes" anywhere in questions A03 to A17 or if questions A01, A18, B01, B03 and C01 are reported then convert all blanks in A03 to A17 to "No". The misinterpretation of the blanks is a potential source of non-sampling error.

The "mark one only" type questions in the paper questionnaire caused processing difficulties because some respondents treated them as "Mark all that apply" type questions and reported multiple responses. In general, the processing system converted multiple responses to "mark one only" questions into "Not stated" responses. Questions A18, E10 and E12 had the highest rate of multiple responses with 5.8%, 8.6% and 7.7% respectively.

Some respondent data for the paper questionnaire did not respect the questionnaire flow instructions. In general, the processing system corrected flow inconsistencies using a top-down approach: the variables that come first in the questionnaire were assumed correct and the other variables were modified accordingly.

8.2.4   Non-response

Total non-response can be a major source of non-sampling error in many surveys, depending on the degree to which respondents and non-respondents differ with respect to the characteristics of interest. Total non-response occurred because the interviewer was either unable to contact the respondent, the respondent refused to participate in the survey or the respondent did not provide sufficient useable data. Total non-response was handled by adjusting the weight of households or individuals 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. Partial non-response is indicated by codes on the microdata file (i.e. Refused, Don't know). As mentioned in Section 7.3, donor imputation was performed to impute the census metropolitan areas for 6% of records.

8.2.5   Measurement of Sampling Error

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

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

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

For example, suppose that, based upon the survey results, one estimates that 39.0% of people in Manitoba travelled to Ontario in the past two years, and this estimate is found to have a standard error of 0.0155. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0155}{0.39} \right) X \ 100 \ \% \ = \ 4.0 \%$$

There is more information on the calculation of coefficients of variation in Chapter 10.0.

9.0   Guidelines for Tabulation, Analysis and Release

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

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 <u>the nearest hundred units</u> 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 <u>to be rounded themselves to one decimal</u> using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

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

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

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

9.2   Sample Weighting Guidelines for Tabulation

The sample design used for the Travel Activities and Motivation Survey (TAMS) 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.

9.3 Categorical Estimates

Before discussing how the TAMS data can be tabulated and analyzed, it is useful to describe the main estimates of population characteristics which can be generated from the microdata file for the TAMS.

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 people in Manitoba who travelled to Ontario in the past two years at the time of the survey, or the proportion of Manitobans who travelled to Ontario, 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. The vast majority of TAMS questions were categorical and, if they were not, they have been grouped so that they have become categorical variables.

> Examples of Categorical Questions
>
> Q: Have you taken any out-of-town trips of one or more nights away from home, for any purpose, in the past 2 years? Include overnight trips to a cottage, cabin or vacation home owned by you or a friend or relative.
> R: Yes / No
>
> Q: How many out-of-town pleasure or vacation trips of one or more nights have you taken in the past 2 years?
> R: None / One / Two / Three / Four / Five or more

9.3.1 Tabulation of Categorical Estimates

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

a)   summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),

b)   summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then

c)   dividing estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.4 Guidelines for Statistical Analysis

The TAMS is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures may differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Approximate variances for

simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

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

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

1) select all respondents from the file who reported RESPSEX = men;

2) 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 RESPSEX = men;

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

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

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

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

9.5   Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the TAMS, users should first determine the quality level of the estimate. The quality levels are *acceptable, marginal* and *unacceptable.* Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 8.0 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to rounded weighted estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	Guidelines
1) Acceptable	Estimates have a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter 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: "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."

9.6 Release Cut-off's for the Travel Activities and Motivation Survey

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 20,000 people possessing a given characteristic in the Atlantic Provinces is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explains the correct procedure to be used for ratios.

Region	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Atlantic Provinces	54,100	& over	13,600	to <	54,100	under	13,600
Québec	81,300	& over	20,200	to <	81,300	under	20,200
Ontario	81,400	& over	20,100	to <	81,400	under	20,100
Manitoba	33,700	& over	8,500	to <	33,700	under	8,500
Saskatchewan	22,100	& over	5,600	to <	22,100	under	5,600
Alberta	93,000	& over	23,500	to <	93,000	under	23,500
British Columbia	125,400	& over	31,700	to <	125,400	under	31,700
Canada	**83,700**	**& over**	**20,600**	**to <**	**83,700**	**under**	**20,600**

Census Metropolitan Areas	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Halifax	24,400	& over	6,400	to <	24,400	under	6,400
Other Atlantic	60,000	& over	15,200	to <	60,000	under	15,200
Quebec City	37,600	& over	9,700	to <	37,600	under	9,700
Montreal	101,400	& over	25,600	to <	101,400	under	25,600
Gatineau	17,600	& over	4,600	to <	17,600	under	4,600
Other Quebec	67,800	& over	17,000	to <	67,800	under	17,000
Ottawa	36,000	& over	9,200	to <	36,000	under	9,200
Kingston	31,800	& over	9,800	to <	31,800	under	9,800
Oshawa	56,600	& over	16,600	to <	56,600	under	16,600
Toronto	111,200	& over	27,900	to <	111,200	under	27,900
Hamilton	46,600	& over	12,200	to <	46,600	under	12,200
St.Catharines-Niagara	47,800	& over	13,300	to <	47,800	under	13,300
Kitchener	33,300	& over	8,800	to <	33,300	under	8,800
London	28,900	& over	7,600	to <	28,900	under	7,600
Windsor	72,900	& over	22,700	to <	72,900	under	22,700
Greater Sudbury	21,300	& over	6,000	to <	21,300	under	6,000
Thunder Bay	24,000	& over	7,200	to <	24,000	under	7,200
Other Ontario	85,100	& over	21,500	to <	85,100	under	21,500
Winnipeg	35,800	& over	9,300	to <	35,800	under	9,300
Other Manitoba	28,700	& over	7,600	to <	28,700	under	7,600
Regina	15,700	& over	4,200	to <	15,700	under	4,200
Saskatoon	19,400	& over	5,200	to <	19,400	under	5,200
Other Saskatchewan	26,600	& over	6,900	to <	26,600	under	6,900
Calgary	83,600	& over	22,200	to <	83,600	under	22,200
Edmonton	83,000	& over	22,200	to <	83,000	under	22,200
Other Alberta	97,500	& over	26,200	to <	97,500	under	26,200
Vancouver	155,100	& over	40,800	to <	155,100	under	40,800
Victoria	52,500	& over	15,200	to <	52,500	under	15,200
Other British Columbia	93,400	& over	24,300	to <	93,400	under	24,300
Canada	**83,700**	**& over**	**20,600**	**to <**	**83,700**	**under**	**20,600**

10.0 Approximate Sampling Variability Tables

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

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

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province which were used to produce the Approximate Sampling Variability Tables for the Travel Activities and Motivation Survey (TAMS).

Region	Design Effect	Sample Size	Population
Atlantic Provinces	2.07	2,481	1,822,494
Québec	2.57	6,794	5,940,869
Ontario	2.44	10,545	9,671,592
Manitoba	2.35	2,067	843,107
Saskatchewan	2.25	2,548	706,325
Alberta	3.32	3,108	2,465,540
British Columbia	4.44	4,156	3,326,176
Canada	**2.93**	**31,699**	**24,776,103**

Census Metropolitan Areas	Design Effect	Sample Size	Population
Halifax	2.01	825	297,463
Other Atlantic	1.85	1,656	1,525,031
Quebec City	2.51	1,276	559,702
Montreal	2.10	2,104	2,868,546
Gatineau	3.04	1,286	222,059
Other Quebec	1.77	2,128	2,290,562
Ottawa	1.92	1,239	668,949
Kingston	2.77	272	117,138
Oshawa	1.76	237	264,517
Toronto	2.43	3,234	4,145,248
Hamilton	2.38	951	554,599
St.Catharines-Niagara	1.58	311	304,046
Kitchener	2.16	763	354,307
London	2.11	877	356,789
Windsor	2.57	240	258,365
Greater Sudbury	1.64	288	123,661
Thunder Bay	1.82	204	97,506
Other Ontario	1.91	1,929	2,426,467
Winnipeg	2.23	1,137	534,034
Other Manitoba	2.60	930	309,073
Regina	2.63	831	151,156
Saskatoon	2.85	853	177,889
Other Saskatchewan	1.79	864	377,280
Calgary	3.04	981	818,632
Edmonton	3.48	1,064	775,066
Other Alberta	3.65	1,063	871,842
Vancouver	4.49	1,709	1,764,224
Victoria	7.18*	1,019	256,193
Other British Columbia	3.00	1,428	1,305,759
Canada	**2.93**	**31,699**	**24,776,103**

* The design effect for the Victoria CMA (census metropolitan area) is high because the response rate in the stratum for telephone numbers of "Unknown" status was particularly low: out of 480 telephone numbers in the initial Random Digit Dialling sample, there were only 21 respondents.

All coefficients of variation in the Approximate Sampling Variability Tables are <u>approximate</u> and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics

Canada on a cost-recovery basis. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a *marginal* estimate could become *acceptable* based on the exact CV calculation.

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

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

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

Rule 1:   Estimates of Numbers of 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 geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2:   Estimates of Proportions or Percentages of 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 <u>proportion</u> of travellers to British Columbia is more reliable than the estimated <u>number</u> of travellers to British Columbia. (Note that in the tables the coefficients of variation decline in value reading from left to right).

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

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular 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 travellers and the numerator is the number of travellers to Quebec.

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

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

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

Rule 5:   Estimates of Differences of Ratios

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

10.1.1 *Examples of Using the Coefficient of Variation Tables for Categorical Estimates*

The following examples based on the 2006 TAMS are included to assist users in applying the foregoing rules.

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

Suppose that a user estimates that 1,919,960 Quebeckers travelled to Ontario in the past two years. How does the user determine the coefficient of variation of this estimate?

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

2)   The estimated aggregate (1,919,960) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 2,000,000.

3)   The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.7%.

4)   So the approximate coefficient of variation of the estimate is 2.7%. The finding that there were 1,919,960 (to be rounded according to the rounding guidelines in Section 9.1) Quebeckers who travelled to Ontario in the past two years is publishable with no qualifications.

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

Suppose that the user estimates that 16,666 / 48,100 = 34.6% of female Haligonians, 55 years and older normally watch sports / sports shows on television. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., female Haligonians 55 years and older), it is necessary to use both the percentage (34.6%) and the numerator portion of the percentage (16,666) in determining the coefficient of variation.

3) The numerator, 16,666, 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. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 35.0%.

4) The figure at the intersection of the row and column used, namely 16.6% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 16.6%. The finding that 34.6% of female Haligonians 55 years and older normally watch sports / sports shows on television is considered marginal. The estimate should be flagged with the letter M (or some similar identifier), and accompanied by a warning to caution subsequent users about the high level of error associated with the estimate.

Example 3:   Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 152,841 / 282,871 = 54.0% of male Edmontonians who have travelled in the past two years said there are many good reasons to travel to Hawaii, while the percentage was 234,191 / 377,196 = 62.1% for female Edmontonians. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the EDMONTON coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for men as 9.2%, and the CV of the estimate for women as 5.5%.

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 (men), $\hat{X}_2$ is estimate 2 (women), 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} = 0.540 - 0.621 = $ -0.081 is:

$$\sigma_{\hat{d}} = \sqrt{[(0.540)(0.092)]^2 + [(0.621)(0.055)]^2}$$
$$= \sqrt{(0.002468) + (0.001167)}$$
$$= 0.0603$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.0603 / 0.081 = 0.744$.

4) So the approximate coefficient of variation of the difference between the estimates is 74.4%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter 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 1,207,129 females read business, finance and investing magazines in a typical month, while 4,541,127 females read fashion and beauty magazines in a typical month. The user is interested in comparing the estimates in the form of the 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 females who read business, finance and investing magazines in a typical month. The denominator of the estimate ($\hat{X}_2$) is the number of females who read fashion and beauty magazines in a typical month.

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

3) The numerator of this ratio estimate is 1,207,129. The figure closest to it is 1,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 4.7%.

4) The denominator of this ratio estimate is 4,541,127. The figure closest to it is 5,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 1.9%.

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

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

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

$$\alpha_{\hat{R}} = \sqrt{(0.047)^2 + (0.019)^2}$$
$$= \sqrt{0.00221 + 0.00036}$$
$$= 0.051$$

6) The obtained ratio of females who read business, finance and investing magazines in a typical month versus females who read fashion and beauty magazines in a typical month is 1,207,129 / 4,541,127 which is 0.266 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 5.1%, which makes the estimate releasable with no qualifications.

Example 5:   Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of females who read business, finance and investing magazines in a typical month versus females who read fashion and beauty magazines in a typical month is 1,207,129 / 4,541,127 = 0.266. Suppose that the user estimates that the same ratio for males is 2,443,090 / 929,580 = 2.63. 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 female ratio ($\hat{R}_1$) and the male ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the female ratio is 5.1%, and 5.7% for the male 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 $\hat{d}$ = 0.266 – 2.63 = -2.364 is:

$$\sigma_{\hat{d}} = \sqrt{\left[(0.266)(0.051)\right]^2 + \left[(2.63)(0.057)\right]^2}$$
$$= \sqrt{(0.000184) + (0.022473)}$$
$$= 0.151$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.151 / (-2.364) = -0.064.

4) So the approximate coefficient of variation of the difference between the estimates is 6.4%, which makes the estimate releasable with no qualifications.

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

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

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

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

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

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

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

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

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

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

10.2.1   *Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits*

A 95% confidence interval for the estimated proportion of female Haligonians 55 years and older who normally watch sports / sports shows on television (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ =   34.6% (or expressed as a proportion 0.346)

t =   2

$\alpha_{\hat{x}}$ =   16.6% (0.166 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

> $CI_{\hat{x}}$ = {0.346 − (2) (0.346) (0.166), 0.346 + (2) (0.346) (0.166)}
>
> $CI_{\hat{x}}$ = {0.346 − 0.115, 0.346 + 0.115}
>
> $CI_{\hat{x}}$ = {0.231, 0.461}

With 95% confidence it can be said that between 23.1% and 46.1% of female Haligonians 55 years and older normally watch sports / sports shows on television.

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

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

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

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

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

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of male Edmontonians who said there are many good reasons to travel to Hawaii, and the proportion for females. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.0603. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.540 - 0.621}{0.0603} = \frac{-0.081}{0.0603} = -1.34$$

Since t = -1.34 is between 2 and -2, no conclusion about the difference between the two estimates is justified 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 TAMS 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. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative

estimate not releasable). Estimates of the variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis.

10.5 Coefficient of Variation Tables

See TAMS2006_CVTabsE.pdf for the coefficient of variation tables for the Travel Activities and Motivation Survey, 2006.

11.0 Weighting

This chapter outlines the derivation of the survey weights for the Travel Activities and Motivation Survey (TAMS). The weighting was done first for the telephone survey, and then for the mail-out/mail-back paper questionnaire.

11.1 Weighting Procedures for the Telephone Survey

1. Calculate design weights

Each of the 132,065 telephone numbers in the sample was assigned a design weight, W_1, equal to the inverse of its probability of selection, calculated as follows within each stratum:

$$W_1 = \left(\frac{\text{Total number of possible sampled telephone numbers}}{\text{Number of sampled telephone numbers}} \right)$$

2. Remove the screened out telephone numbers

Telephone numbers identified as out-of-service numbers prior to data collection were dropped from weighting. There were 13,550 such numbers.

3. Adjust for non-resolved telephone numbers

There were 9,140 telephone numbers that were not resolved during data collection, i.e. telephone numbers which were not determined as in-scope (residential) or out-of-scope (business or non-working number). The weights were adjusted as follows within each province by stratum:

$$W_2 = \left(\frac{\sum W_1 \text{ for resolved telephone numbers} + \sum W_1 \text{ for unresolved telephone numbers}}{\sum W_1 \text{ for resolved telephone numbers}} \right) \times W_1$$

4. Remove out-of-scope telephone numbers

Telephone numbers identified as businesses, out-of-service numbers, or out-of-scope numbers, such as cottage telephone numbers, were dropped. There were 19,341 such telephone numbers.

5. Adjust for non-responding households

There were 31,166 records which were confirmed as in-scope (residential) but the household roster listing all the members of the household was not completed. The weights were adjusted as follows within each province by stratum:

$$W_3 = \left(\frac{\sum W_2 \text{ for responding households} + \sum W_2 \text{ for non - responding households}}{\sum W_2 \text{ for responding households}} \right) \times W_2$$

6. Adjust for the selection of one household member

During collection, one member of the household aged 18 years of age or older was randomly selected to participate in the survey. The weights were calculated as follows:

$$W_4 = \text{Number of eligible household members} \times W_3$$

7. Adjust for non-responding persons

There were 5,311 records with a complete household roster, but the person selected to participate in the survey did not respond (i.e. data for the first question, TS_Q01, is not reported). Propensity to respond was modelled using a logistic regression model within each region (Atlantic Provinces, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia), with explanatory variables sex and age group (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over) of the selected respondent, household size (one, two, three, four or more), presence of children in the household (Yes / No), initial telephone status (residential / unknown) and language of interview (English / French). Non-response groups were formed within each region based on the propensity to respond. The weights were adjusted as follows within each non-response group:

$$W_5 = \left(\frac{\sum W_4 \text{ for responding persons} + \sum W_4 \text{ for non-responding persons}}{\sum W_4 \text{ for responding persons}} \right) \times W_4$$

8. Adjust for records with insufficient data

There were 407 records with some data, but not a sufficient amount to be considered useable. Adjustment factors were calculated within each region (Atlantic Provinces, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia) by TS_Q01 in order to preserve the distribution of the number of travellers. The weights were adjusted as follows within each weighting group:

$$W_6 = \left(\frac{\sum W_5 \text{ for records with sufficient data} + \sum W_5 \text{ for records without sufficient data}}{\sum W_5 \text{ for records with sufficient data}} \right) \times W_5$$

9. Adjust for number of telephone lines

Weights for households with more than one telephone line (with different telephone numbers) were adjusted downwards to account for the fact that such households have a higher probability of being selected. The weight of each respondent was divided by the number of distinct residential telephone lines (up to a maximum of 4) that serviced the household. If the number of lines was missing, a value of one was imputed. The weights were calculated as follows:

$$W_7 = \left(\frac{W_6}{\text{Number of telephone lines in the household}} \right)$$

10. Calibrate to external totals

Calibration was performed using February 2006 demographic counts. Two sets of demographic counts were used: counts of persons aged 18 and over at the census metropolitan area (CMA) level, and age group (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over) by sex counts at the region (Atlantic Provinces, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia) level. Generalized regression (GREG) estimation was used to calibrate the weights so that the sum of the weights equal the demographic counts in both dimensions. The calibrated weights are denoted W_8.

The weights of the 7,007 non-travellers were derived using Steps 1 to 10.

11.2 Weighting Procedures for the Mail Survey

A separate set of weights were derived for the mail-out/mail-back paper questionnaire. The data from the telephone survey was used to determine which variables best explain non-response to the paper questionnaire. The following steps were performed:

11. Adjust for non-response to paper questionnaire

Non-response weighting adjustments were performed to take into account the 21,451 travellers with no paper questionnaire. Adjustment factors were calculated within each region (Atlantic Provinces, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia) by age group (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over) by sex by level of education (High school or less, University degree, Other post-secondary). Level of education was dropped for cells with fewer than 20 respondents to the paper questionnaire or fewer than 30 respondents to the telephone interview. The weights were adjusted as follows within each weighting group:

$$W_9 = \left(\frac{\sum W_7 \text{ for all telephone travellers}}{\sum W_7 \text{ for travellers with complete questionnaire}} \right) \times W_7$$

12. Calibrate to telephone survey

The telephone data of the 46,143 travellers who completed the telephone interview were used to calculate control totals using weight W_8 from weighting Step 10. The following sets of control totals were calculated (at the region level unless otherwise stated):

- Number of travellers by age group (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over) by sex;
- Number of travellers aged 18 and over at the CMA level;
- Number of travellers born in Canada (based on DM_Q09);
- Number of travellers in households with one, two or three adults or more (from roster);
- Number of persons who travelled within own province (TS_Q02A);
- Number of persons who travelled to other Canadian provinces / territories (TS_Q02B);
- Number of persons who travelled to the United States (TS_Q02C);
- Number of persons who travelled to another location (TS_Q02D to TS_Q02K).

The weights of the 24,692 mail-out respondents were calibrated so that they produce the same estimates as these control totals (again using GREG estimation). Records with missing data in the calibration variables were temporarily imputed. The calibrated weights are denoted W_{10}.

Note that the weights were calibrated so that, for example, estimates of TS_Q02C (not A01_Q14A) using the 24,692 mail-out respondents and weights W_{10}, are equal to the estimates of TS_Q02C using the 46,143 telephone travellers and weights W_8. Questions A01_Q01A to A01_Q21B from the paper questionnaire were not used in the calibration because there were too many discrepancies between the data from the telephone interview and the paper questionnaire, and using the two sources of data could have caused biases in the weights.

The weights of the 24,692 travellers were derived using Steps 1 to 12.

12.0 Questionnaires

Refer to the files identified below for the questionnaires for the Travel Activities and Motivation Survey (TAMS) microdata:

TAMS2006_M_QuestE.pdf   (Mail questionnaire)

TAMS2006_T_QuestE.pdf   (Telephone questionnaire)

13.0 Record Layout with Univariate Frequencies

See TAMS2006_CdBk.pdf for the record layout with univariate counts.