TRAVEL ACTIVITIES AND MOTIVATION SURVEY, 1999

USER GUIDE

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Travel Activities and Motivation Survey (TAMS) was conducted by Statistics Canada on behalf of the Canadian Tourism Commission and four provincial and territorial agencies responsible for tourism.

The survey was conducted under the voluntary provisions of the Statistics Act, Revised Statutes of Canada, 1985, Chapter S19. Collection plans for the survey are registered under collection registration number STC/SSD-040-75154 and personal information bank number STC/P-PU-016.

This documentation manual contains information to access and manipulate data from the survey.

Any questions 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

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

2.0 Background

The types of information collected are: areas of Canada travelled to in the previous two years and travel intentions for the next two years; reasons non-travellers do not travel; participation in recreational and entertainment activities; reasons for travelling in Canada and to Canadian provinces and territories; types of accommodation used while travelling; sources of travel planning information; and impressions of parts of Canada as travel destinations.

Data from the survey were collected under the data-sharing provisions of the Statistics Act. Agreements to share the data collected from the survey were signed with the following departments and agencies:

1. Canadian Tourism Commission;
2. Ontario Ministry of Economic Development, Trade and Tourism;
3. Saskatchewan Tourism Authority;
4. Manitoba Ministry of Industry, Trade and Tourism; and
5. Yukon Government Department of Tourism.

Data from the Travel Activities and Motivation Survey are used by a number of federal government agencies and departments, the Canadian Tourism Commission and provincial tourism boards. Other users include the media, business, consultants, universities and other researchers interested in the Canadian traveller.

Researchers and consultants in the government, private businesses, universities and the media will use the survey results to educate and inform the public, develop new programs and determine the need for new services and infrastructure.

3.0 Objectives

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

Some of the additional data objectives were:

- to collect information on overnight trips taken in the past two years in Canada, the USA and other countries;

- to collect information on the reasons for overnight trips taken in Canada, the USA and other countries;

- to collect information on activities undertaken while travelling; and

- to collect information on motivation to take trips in the future.

4.0 Concepts and Definitions

This chapter outlines the concepts and definitions of interest to users. The Travel Activities and Motivation Survey (TAMS) computer-assisted telephone interviewing (CATI) application was made up of two components for data collection; the control component and the questionnaire component.

The control component was used a) to maintain a record of calls for each telephone number in the sample, b) to list the household members as well as their age and gender and c) select a respondent to complete the TAMS telephone interview questionnaire portion.

The TAMS questionnaire was completed by each selected respondent, and was made up of two content components, one for the telephone interview survey and one for the mail survey, if the respondent indicated in the telephone interview that he/she a) travelled in the past two years or b) expected to travel in the next two years.

The concepts and definitions for the TAMS control component are described in Section 4.1 while those specific to the TAMS questionnaire component are provided in Section 4.2.

Users are referred to Chapter 12.0 of this document for a copy of the actual survey questions.

4.1 Control Component Concepts and Definitions

Dwelling
A dwelling is a set of living quarters in which a person or a group of people reside or could reside (e.g. a private home, an apartment, a duplex).

Household
A household is any person or group of people living in a dwelling.

For example:

- one person living alone;
- a group of people who are not related, but who share the same dwelling;
- a family.

Household respondent
The household respondent is the person who provides general information about the household member(s).

Household members
A household member is any person who, during the survey period:

- regards or is reported to regard the dwelling as his/her usual place of residence; or
- is staying in the dwelling and has no usual place of residence elsewhere.

This includes:

- a spouse or partner (common-law or same-sex) who usually resides in the dwelling but may be away temporarily due to work or school, for example a member of the Canadian Armed Forces may be posted to another region but this dwelling is his/her usual place of residence;
- a child temporarily away from home due to school or seasonal work but who consider 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 a joint "physical" custody situation if the children spend substantial portions of time with each parent and the children slept-over in this dwelling the night before;
- 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 student visas;
- a person staying in Canada on work permits and his family.

Selected person (respondent)
The selected person (respondent) is the household member, 18 years of age or older, who is randomly chosen to provide the information for the questionnaire.

In scope
Telephone numbers are considered in scope if they are servicing:

- private dwellings; or
- private quarters within collective dwellings.

Collective dwellings would include rooming houses, student residences, hostels, religious residences, military camps, etc.

Out of scope
Telephone numbers are considered out of scope if the telephone number services:

- an individual residing in an institution, for example, nursing home, hospital, correctional facility, etc.; or
- a collective dwelling for which the telephone number is used by 10 or more unrelated individuals.

4.2 Questionnaire Component Concepts and Definitions

Definition of an Overnight Trip
Any trip of one night or more regardless of distance from home. What constitutes a "trip" is left to the interpretation and perception of the respondent. For example, a sleep over at a friend's house next door, going to the cottage for the week-end, staying at your parents' place while your apartment is being painted, are considered trips if the respondent perceives this as such.

The following are not trips:

- taking an ambulance to a hospital or clinic;
- travelling as an employed operator or vehicle crew member (e.g., truck driver, flight attendant, boat captain);
- commuting to school or a usual place of work;
- actually moving from one residence to another (Note that home-hunting before the actual move is considered a trip.);
- trips that began when a respondent was not a resident of Canada (Trips must have originated in Canada.);
- trips longer than a year or 365 days.

Main reason for the trip
This question determines why a selected respondent travelled.

If another person went on the trip with the selected respondent, record only the main reason the selected respondent went on the trip.

Pleasure Trip
Include trips taken for shopping, sightseeing, accompanying someone on a business trip or attending a sporting event, a concert, a movie, or music festival.

Include trips where the main reason for the trip was to shop around or to make purchases that were not available closer to home.

In some areas of Canada, respondents may take short overnight trips to the United States to shop. It is important that all of these trips are reported under "Pleasure".

Visiting friends or relatives
The main motivation for the trip was to visit friends or relatives even though the selected respondent may have participated in other activities while on the trip.

Business
The main purpose of the selected respondent's travel is to carry out the duties of employment. Business trips include meetings, conferences, seminars, training, inspecting, consulting, or attending a convention.

Other personal reasons
The main motivation for the trip is to keep an appointment with a physician, a lawyer, a dentist; to go to a funeral; or a job interview, etc.

Include trips caused by emergency evacuation.

A trip associated with a church function will be reported according to the type of event attended. The interviewer should probe respondents to determine if the purpose was personal, such as a retreat, or if the purpose was for pleasure, such as a social event organized by the church.

Income
The respondent should report income from all sources. For example:

- from wages and salaries or self-employment, including tips, commissions and gratuities;
- from government sources such as employment insurance, Canada or Quebec pension plan, old age security;
- from interest, dividends, investments or private pensions;
- from other sources, such as alimony, scholarship.

Include the total amount from all jobs before deductions for income tax, pension contributions, health insurance, etc.

Exclude gambling gains, lottery winnings, money inherited during the year in a lump sum, income tax refunds, loan payments, lump sum settlements of insurance policies, rebates received from property taxes, pension contribution refunds, strike and sick pay from trade unions, etc.

Level of education
This refers to the type of high school education certificates, diplomas, degrees or other types of post-secondary education.

Elementary/ secondary education
The responsibility for education in Canada rests with provincial and territorial governments. Each province and territory has developed its own system for education, and the structure can differ from jurisdiction to jurisdiction.

The following table illustrates the similarities and differences as of 1997-1998.

Province/ Territory	Number of Years Pre-grade 1	Primary/ Elementary (by grade)	Junior High/ Intermediate/ Middle (by grade)	Senior High (by grade)	Secondary/ High School (by grade)
Newfoundland and Labrador	1	1-6	7-9	10-12	
Prince Edward Island		1-6	7-9	10-12	
Nova Scotia	1	1-6	7-9	10-12	
New Brunswick (english sector)	1	1-5	6-8	9-12	
New Brunswick (french sector)	1	1-8			9-12
Quebec	2	1-6		Secondary 1-5	
Ontario	2	1-8			9-12 (Includes Ontario Academic Course)
Manitoba	1	1-8			9-12
Saskatchewan	3	1-5	6-9		10-12
Alberta	1	1-6	7-9	10-12	
British Columbia	1	1-7			8-12
Yukon	1	1-7			8-12
Northwest Territories or Nunavut	1	1-6	7-9	10-12	

Source: Education Indicators in Canada, 1998-1999, Catalogue no. 81-582-XPE.

Trade/ vocational certificate or diploma
This term is used to classify skill courses that prepare trainees for occupations not at the professional or semi-professional levels. A trade-vocational school is a public educational institution which offers courses to prepare people for employment in a specific occupation such as a heavy equipment operator, automotive mechanic and upholstering. Many community colleges or technical institutes offer certificates or diplomas at the trade level.

Community college, technical college (collège d'enseignment général et professionnel (CEGEP) in Quebec) certificate or diploma
Community college or CEGEP certificates or diplomas are usually offered after one to two years of study.

NOTE: 1. Ryerson Politechnical Institute and many B.C. colleges award both university and college-level diplomas and certificates.

2. The following are equivalences in the former systems of education:

Bible colleges/seminary = baccalaureat (Priesthood) theology
Nursing (infirmière) = community college diploma

Other education
This includes:

- diplomas, certificates or licenses obtained through professional associations such as in accounting, banking or insurance;
- non-professional health certificates (i.e. CPR, First Aid, Red Cross); and
- continuing education, personal interest courses.

5.0 Survey Methodology

The telephone survey was carried out from mid-October to mid-December 1999 using a Random Digit Dialling (RDD) telephone sampling method. The mail survey, conducted between November 1999 and March 2000, used the addresses obtained during the telephone survey to contact respondents.

Because the mail survey is a complement to the telephone survey, they have many commonalities in the survey design. In the documentation that follows, we begin each sub-section by talking about the telephone survey and then follow it up with what changes for the mail survey.

5.1 Population Coverage

The target population for the telephone survey was all persons 18 years of age and older in each of the ten Canadian provinces, excluding full-time residents of institutions. With RDD, households without telephones were also excluded. However, persons living in such households represented less than 2% of the target population. Survey estimates have been adjusted through weighting to represent persons without telephones.

The target population for the mail survey was the same as just described except that it excluded non-travellers. Travellers were defined as persons answering, at the time of the telephone survey, that they had taken a trip in the past two years or were very likely or fairly likely to take a trip in the next two years.

5.2 Elimination of Non-working Banks Random Digit Dialling Design

Random digit dialing is a method of obtaining respondent data through the random calling of households. The method we use involves the concept of banks. As an example, consider the phone number, including area code, of 613-951-9999. The first six numbers form the prefix and the first eight numbers (italicized) form the bank. The two sources of phone number information we have are phone billing files sold to us by the phone companies and an electronic version of all phone numbers appearing in paper directories covering all provinces and territories. For unlisted numbers, the billing file typically gives us the bank, excluding the last two numbers for confidentiality. Cell phone numbers are avoided on these lists because the prefixes allocated for cellular phones by the telephone companies are generally dedicated prefixes for cell phone use only. We use these two sources to identify non- empty banks, the basis of our RDD frame. This method is called the Elimination of Non-Working Banks. Each bank was assigned to a stratum within its province.

A random sample of telephone numbers was generated in each stratum (from the working banks). An attempt was made to predict the percentage of numbers dialed that would reach a household (this is known as the “hit rate”). The hit rate was estimated using information from previous RDD surveys, specifically the most recent cycle of the General Social Survey. For the Travel Activities and Motivation Survey (TAMS), we anticipated a hit rate of 43%. The actual hit rate was 49%.

Once a complete list of household members was created, we identified the eligible sample members and selected one at random for the telephone interview. If the selected person completed the telephone interview and was a traveller, we asked them to complete a mail questionnaire.

5.3 Sample Design

The sample design for the telephone survey is a stratified simple random sample of telephone numbers. Generally, the strata represented the Census Metropolitan Areas (CMA) of the province and the non-CMA areas. The rationale in stratifying by sub-provincial areas was to be able to break down final estimates by stratum.

The sample design for the mail survey is a two phase design, where the first phase is the telephone survey. We use the telephone survey data to define the second phase persons as travellers. We take a census of all second phase travellers.

The sample design is a complex design because our target population is defined as adults but the sampling unit is telephone numbers. It is at the weighting stage where we calculate the initial probability of selection of the people in the sample, which is based on the initial telephone probability of selection, the hit rate, the number of telephones in a household and the number of adults in the household.

5.4 Sample Sizes

The sample consisted of 60,000 telephone numbers nationally. Upon contacting a household, all household members were listed and basic demographic information on the age and sex of each household member was collected. From this list, we randomly selected a person 18 years of age or older and conducted a telephone interview. A total of 29,123 people were eligible for the telephone survey and 18,385 of them completed it.

The travellers who completed the survey were asked to accept and complete the mail questionnaire. From a total sample size of 24,769 eligible travellers, we obtained 5,740 completed mail questionnaires.

The table in Section 8.1 provides a more detailed breakout of the sample sizes and response rates by stratum.

6.0 Data Collection

The survey was a Random Digit Dialling (RDD) sample survey of 18,385 Canadians and consisted of two phases: a telephone survey to identify travellers and non-travellers, and a mail survey, which was completed by all travellers.

The telephone survey was conducted using a computer-assisted telephone interviewing (CATI) application known as CASES (computer-assisted survey execution system). Data collection for the telephone survey took place for approximately a ten week period from mid-October to mid-December, 1999. The CATI instrument was administered through the Statistics Canada Regional Offices in Montreal, Sturgeon Falls, Winnipeg and Vancouver. In most offices, telephoning was conducted in two shifts, between 9:00 a.m. and 9:00 p.m.

All regional office staff who worked on the survey (i.e. project supervisors, senior interviewers and interviewers) were given a training session of 1 to 1-1/2 days designed to familiarize them with the purpose and concepts of the survey, the CATI questionnaire and procedures involved and some basic techniques of telephone interviewing. A Travel Activities and Motivation Survey (TAMS) Content/Interviewer's Manual and printed copies of the TAMS questionnaires were provided as support documentation for interviewers during the data collection period. Most of the interviewers had previous telephone interviewing experience.

Respondents were approached in the principal official language of their choice. Bilingual interviewers were used where required.

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

After all attempts to interview a selected respondent were exhausted, interviewers coded the final result of the interview on the CATI questionnaire record, using one of the final status codes. See the table in Section 8.1 for more details.

Data collection for the mail survey took place for approximately a five month period from November, 1999 to March 2000. The mail survey was also administered through the Statistics Canada Regional Offices in Montreal, Sturgeon Falls, Winnipeg and Vancouver. In most offices, a telephone follow-up was conducted in January and February, 2000 between 9:00 a.m. and 9:00 p.m.

7.0 Data Processing

7.1 Data Capture

The CASES (computer-assisted survey execution system) system has two main parts; Case Management and a component that is specific to each survey.

The Case Management system controls the case assignment and data transmission for the survey. For the Travel Activities and Motivation Survey (TAMS), a case referred to an individual respondent selected from the Random Digit Dialling (RDD) sample. The Case Management system also automatically recorded 'survey management' information for each contact (or attempted contact) with respondents, and provided reports on the progress of the collection process throughout the entire interview period.

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

As cases were completed, data files containing the records of all captured information were electronically transmitted to Ottawa. This resulted in the creation of a raw data file to be used in post-collection survey processing.

All the mail survey questionnaires completed by respondents were mailed back to the four Regional Offices. From there they were shipped in batches to Operations and Integration Division in Head Office (O.& I.D.) in Ottawa for data capture.

Capture of the mail survey questionnaire was accomplished using desktop computers. During this process any document containing at least a CASEID was captured. An unedited version of all records was electronically transmitted to Special Surveys Division for post-collection survey processing.

7.2 Edit and Imputation

The raw survey data coming out of the TAMS CASES application was generally cleaner than survey data collected using a paper-and-pencil telephone interview collection method. Nevertheless, some post-collection editing was still necessary to ensure that users of the final microdata file could clearly distinguish between valid responses, item non-response and valid skips in survey questions.

As part of the overall editing strategy, a set of codes which could be applied to every question to describe the various types of 'non-answers' was developed. The coding strategy currently being used by many surveys, including the TAMS, is as follows:

- 6, 96, 996, etc. for valid skips (ie., question does not apply to respondent)
- 7, 97, 997, etc. for 'don't know' response to question
- 8, 98, 998, etc. for 'refused' response to question
- 9, 99, 999, etc. for not stated (ie., respondent may or may not belong to target population for question; usually happens if one of the questions defining the target population was answered as 'don't know' or 'refused', or if respondent quit before end of survey)

Preliminary grooming of the CATI\CASES response file included recoding the response values for categorical questions from 'x's to '7's (a 'don't know' response) , 'r's to '8's (a ' refusal'), and blanks to valid skips (6, 96, 996, etc.), as per the coding strategy mentioned above.

As well, all multiple choice questions were reformatted to a series of yes/no questions.

Text was removed where open-ended questions were asked and replaced by character flags. An 'other, specify' entry file was created for further processing during the coding stages. For more details please see Section 7. 3.

Since the flows between questions were already programmed in the CATI application, completed interviews had every on-path question answered. Consequently, it was not possible to skip a question which applied to that respondent. Furthermore, the CATI program ensured that every response entered was within the valid range defined for the question. The deterministic edit was based on these assumptions: if a respondent was supposed to answer a question, then whatever response was present was valid – there was no verification done. If a question did not apply to a respondent, then a 'valid skip' code was assigned. If there was any uncertainty about whether or not a question applied to a respondent, then a 'not stated' code was assigned.

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

In addition all non response records were removed from the file as were records for respondents who did not wish to share their answers with other government departments .

For the mail survey the same standard code set used in the telephone survey was used for not stated and valid skip answers, etc. (as specified earlier in this section). However a different overall approach to editing was needed. The responses (or lack of responses) to questions 1 to 12 were checked to determine the response to certain key questions (Q1a, Q1b, Q2a, Q2b, Q3a, Q8a). For the remaining questions a combination of top-down and bottom up editing was done. Two main types of errors in questionnaire flow were corrected.

The first type of error was where questions that did not apply to the respondent, and should therefore not have been answered, were found to contain answers. In this case the computer edit automatically eliminated superfluous data by assigning the valid skip code.

The second type of error was where questions that should have been answered were left blank. These questions were assigned the not stated code.

In addition, some edits were done to ensure consistency between questions. However no consistency edits were done between the telephone and mail surveys and, unlike the telephone survey, records belonging to non-sharers were retained on the microdata file.

7.3 Coding of Open-ended Questions

A number of data items on each telephone questionnaire were recorded in an open-ended format. Some of these 'other specify' entries were subsequently recoded.

Each text entry was manually checked and a decision made as to whether or not it belonged in one of the existing categories for that question. If it was determined that it did, then the answer to the 'other specify' was changed to 'no' and the text removed. The appropriate existing category was switched to 'yes'. For Q36 new categories were created in addition to recoding some text responses to existing categories.

There were no open-ended format questions on the mail survey.

7.4 Creation of Derived Variables

A number of data items on the telephone microdata file were derived by combining similar items from the questionnaire to create such variables as number of people in the household, number of people in the family, the number of household members under 18 years of age and number of family members under 18 years of age, etc. More complete descriptions are provided for each derived variable in the codebook.

Many of the derived variables on the telephone file were suppressed for reasons of confidentiality. Please see Section 7.6 for further information. Special tabulations are possible based on the master file.

There are no derived variables as such on the mail microdata file. However certain variables such as province, household size, etc. were copied from the telephone survey to make it easier for users of the mail microdata file.

7.5 Weighting

The principle behind estimation in a probability sample 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 sample of 2% of the population, each person in the sample represents 50 people in the population. The weighting phase is a step which calculates, for each record, this number. This survey weight appears on the microdata file and must be used to derive meaningful population estimates from the survey. For example, if the number of people who travelled in the past two years in Canada is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights associated with those records.

Since the telephone and mail surveys have different target populations and response rates, there are separate weights for each data file. 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 described in this document differ in a number of important respects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. For example, variables with extreme values have been capped (e.g. number of persons in a household), information for some variables have been aggregated into broader categories (e.g. age) and selected variables for particular respondents individually suppressed. 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.

8.0 Data Quality

8.1 Response Rates

The following table presents a summary of the number of respondents selected in the telephone and mail surveys for each sub-provincial stratum, and the response rates achieved. The hit rate is the percentage of randomly generated telephone numbers that turned out to be residential lines (as opposed to businesses or non-working numbers). The telephone response rate is the percentage of telephone interviews completed based on the number of residential telephone numbers. The mail response rate is based on the estimated number of travellers with residential telephone numbers. Therefore, the non-response from the telephone survey affected the response rate for the mail survey.

Response Rates by Stratum for the Public Use Microdata Files

Stratum	Telephone Numbers Generated	Hit Rate (%)	Telephone Response Rate (%)	Number of Telephone Responses	Mail-out Response Rate (%)	Number of Mail-out Responses
St. John's CMA	1,662	42.2	64.8	455	23.9	141
Halifax CMA	2,080	42.0	61.3	536	23.7	182
Atlantic Other	2,225	40.2	62.3	557	23.5	163
Montreal CMA	2,361	56.2	56.5	750	19.6	217
Quebec CMA	1,614	56.5	64.3	586	27.6	197
Quebec Other (excluding Hull)	1,458	51.2	67.1	501	26.2	140
Ottawa-Hull CMA	3,460	49.1	60.2	1,023	25.1	364
Toronto CMA	5,787	44.6	67.1	1,730	21.0	444
Hamilton CMA	2,872	50.5	69.2	1,004	26.4	313
London CMA	2,318	51.0	52.4	619	21.9	227
Kitchener CMA	2,390	55.9	52.4	700	21.0	246
St Catharines-Niagara CMA	2,336	50.9	71.2	846	26.8	252
Windsor CMA	2,211	57.2	48.3	610	16.0	178
Oshawa CMA	2,184	58.2	72.2	917	26.9	289
Sudbury CMA	1,969	41.4	60.4	492	22.1	155
Thunder Bay CMA	1,844	54.1	55.5	554	18.7	163
Ontario Other	3,497	45.0	53.5	842	19.6	263
Winnipeg CMA	1,944	47.6	73.4	679	26.4	199
Manitoba Other	1,647	38.6	78.4	498	33.8	175
Saskatoon CMA	1,196	51.8	73.0	452	27.4	156
Regina CMA	1,316	44.4	69.2	404	24.9	129
Saskatchewan Other	1,904	38.4	72.1	527	24.3	152
Edmonton CMA	1,589	54.2	65.3	562	23.7	184
Calgary CMA	1,692	51.4	63.6	553	21.7	175
Alberta Other	1,380	43.8	71.9	435	26.4	144
Vancouver CMA	2,036	51.4	60.4	632	20.2	188
Victoria CMA	1,543	56.3	64.3	559	23.1	183
British Columbia Other	1,166	48.3	64.3	362	23.2	121
Canada	**59,681**	**48.8**	**63.1**	**18,385**	**23.2**	**5,740**

8.2 Survey Errors

The quality of data can be divided into two aspects: sampling error and non-sampling error. The sampling error comes from taking a sample rather than a census and can be quantified using standard statistical tools. The sampling variability tables discussed in Chapter 10.0 are one way to evaluate sampling error.

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. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included focus group testing of the questionnaire, testing of processes through a pilot survey, use of highly skilled interviewers, training of interviewers with respect to the survey procedures and questionnaire, and coding and edit checks to verify the processing logic.

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.

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. Since the non-response to the mail is much higher than the non-response to the telephone survey, we suggest that users, if given the choice, rely more heavily on the estimates coming from the telephone survey.

The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households who responded to the survey to compensate for those who did not respond.

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

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

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results. The standard error of an estimate is often 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, based upon the survey results, suppose that one estimates that 8.3% of people in Ontario travelled to Alberta in the past two years, with a standard error of 0.3%. Then the coefficient of variation of the estimate is calculated as:

$$C.V. = \left(\frac{.003}{.083} \right) x\ 100\% = 3.6\%$$

In the case of Travel Activities and Motivation Survey (TAMS), we can see evidence of non-sampling error through use of a question that was asked on both the telephone and mail questionnaires. In the telephone survey, it was worded as follows:

> Q5. You indicated that you took at least one pleasure trip in Canada in the past two years. In which of the following provinces or territories did you stay (either passing through or as your destination) for one or more nights on these pleasure trips?
>
> (10) Newfoundland and Labrador
> ⋮ ⋮
> (35) Ontario

On the mail survey, the respondent was asked the following question:

> Q16B. Have you taken a pleasure trip of one or more nights in the past 2 years to any location in Northern or Southern Ontario?

The following table shows the response to these questions and the absolute differences in the percentage estimated by both surveys. The startling observation is that the mail survey has a consistently higher estimated percentage. This cannot be attributed to differences in the way the travelling population has been defined. We speculate that this is a case of respondent recall error. It just so happened that the telephone survey was conducted just before Christmas and the mail survey was mostly conducted after Christmas. It is likely that respondents tend to better remember travel in the past few months than trips taking place before then and hence the consistently higher percentage for the mail survey.

As a result of this comparison, the user of the data ought to be aware that the answer received does not necessarily answer the question asked. In short, the information derived from these two questions is not very accurate. This is a case where the sampling error is swamped by the non-sampling error due to problems with respondent recall. Users are thus cautioned that the quality of the data cannot be completely summarized by use of sampling variability estimates alone.

Estimated percentage of travel to Ontario in the past two years among all travellers

Stratum	Telephone Estimate (%)	Mail Estimate (%)	Difference (%)
Toronto CMA	43.8	74.0	30.2
Thunder Bay CMA	52.4	82.3	29.9
Hamilton CMA	55.4	83.6	28.2
Oshawa CMA	55.9	81.1	25.2
St Catharines-Niagara CMA	58.1	83.1	25.0
Winnipeg CMA	32.3	56.8	24.5
Kitchener CMA	61.4	83.8	22.4
London CMA	64.2	84.9	20.7
Ottawa CMA	59.3	79.6	20.3
Ontario Other	67.1	87.3	20.2
Sudbury CMA	76.4	90.0	13.6
Atlantic Other	26.3	39.8	13.5
Windsor CMA	65.9	79.3	13.4
Saskatchewan Other	9.1	22.4	13.3
Alberta Other	10.8	23.1	12.3
Hull CMA	31.3	43.0	11.7
Montreal CMA	28.8	40.3	11.5
Manitoba Other	20.8	32.3	11.5
British Columbia Other	11.8	22.3	10.5
Halifax CMA	31.6	40.4	8.8
Victoria CMA	18.1	26.7	8.6
Calgary CMA	17.5	24.9	7.4
Regina CMA	16.6	23.2	6.6
Quebec Other (excluding Hull)	19.2	25.7	6.5
Saskatoon CMA	13.9	19.8	5.9
Quebec CMA	26.2	31.6	5.4
St. John's CMA	27.2	32.4	5.2
Edmonton CMA	13.6	18.3	4.7
Vancouver CMA	17.5	22.1	4.6
Canada	**34.1**	**49.6**	**15.5**

9.0 Guidelines for Tabulation, Analysis and Release

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata files.

9.1 Rounding Guidelines

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

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

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

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

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the 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.

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

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 that travelled to Ontario in the past two years at the time of the survey, or the proportion of people in Manitoba 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.

Estimates of the number of people with a certain characteristic, e.g. having travelled in the past two years, can be obtained from the microdata file by summing the final weights of all records possessing that characteristic of interest. Details of how to calculate proportions and ratios of the form X/Y are presented in Chapter 10.0.

9.3 Guidelines for Statistical Analysis

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

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Variances for simple estimates such as totals, proportions and ratios (for qualitative variables) are provided in the accompanying Sampling Variability Tables.

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

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

a) select all respondents from the file who reported SEX = male

b) Calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported SEX = male

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

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

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

The calculation of truly meaningful variance estimates requires detailed knowledge of the design of the survey and the adjustments that were made to take into account non-response. Such detail cannot be given in this microdata file because of confidentiality. One way to mitigate the problem is to use the design effects given in the first table of Chapter 10.0. Find the area to which the inference is to be made and then multiply the estimate of variance (usually assumes a simple random sample) by the design effect. This will, in a rough way, adjust for the fact that the sample design includes stratification and that we make adjustments for non-response. Statistics Canada is able to do more exact variance calculations that take the complete sample design into account on a cost recovery basis.

9.4 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the Travel Activities and Motivation Survey, 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 weighted rounded estimates.

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

Quality Level Guidelines

Quality Level of Estimate	Guidelines
Acceptable	Estimates have a sample size of 30 or more and low coefficients of variation, in the range 0.0% to 16.5% No warning is required.
Marginal	Estimates have a sample size of 30 or more and high coefficients of variation, in the range 16.6% to 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
Unacceptable	Estimates have a sample size of less than 30 or very high coefficients of variation, in excess of 33.3%. Statistics Canada recommends that estimates of unacceptable quality not be released. 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."

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from these microdata files and which could be readily accessed by the user, two sets of Approximate Sampling Variability Tables have been produced. These 'look-up' tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation (CV) are derived using the variance formula for simple random sampling and incorporating a factor which reflects the stratified nature of the sample design and the 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 to be used in the look-up tables which would then apply to the entire set of characteristics.

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost recovery basis. The use of actual variance estimates would allow users to release otherwise unreleaseable estimates, i.e. estimates with coefficients of variation in the 'confidential' range.

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

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

Input data for approximate sampling variability tables

Area	Telephone Survey			Mail-out Survey		
	Design Effect	Sample Size	Population Size	Design Effect	Sample Size	Population Size
Canada	**2.21**	**18,385**	**23,261,571**	**2.91**	**5,740**	**19,618,857**
Atlantic Provinces	2.32	1,548	1,829,572	2.77	486	1,455,265
Quebec	1.64	2,103	5,755,306	2.09	640	4,547,941
Ontario	2.39	9,071	8,900,616	3.70	2,808	7,522,137
Manitoba	1.28	1,177	828,107	1.73	374	688,964
Saskatchewan	1.42	1,383	724,612	1.59	437	642,229
Alberta	1.33	1,550	2,177,501	1.76	503	1,980,387
British Columbia	1.82	1,553	3,045,857	2.41	492	2,781,933
St. John's CMA/Atlantic Other	2.05	1,012	1,553,109			
Halifax CMA	1.46	536	276,463			
Atlantic Other	1.34	557	1,418,045			
Montreal CMA	1.29	750	2,694,544			
Quebec CMA	1.53	586	575,614			
Quebec Other (excluding Hull)	1.29	501	2,275,061			
Ottawa-Hull CMA	1.32	1,023	857,198			
Toronto CMA	1.25	1,730	3,608,933			
Hamilton CMA	1.23	1,004	517,168			
London CMA	1.28	619	331,351			
Kitchener CMA	1.28	700	318,984			
St Catharines-Niagara CMA	1.31	846	288,348			
Windsor CMA	1.30	610	228,587			
Oshawa CMA	1.31	917	216,040			
Sudbury/Thunder Bay CMA	1.30	1,046	235,638			
Ontario Other	1.34	842	2,508,455			
Winnipeg CMA	1.29	679	539,559			
Manitoba Other	1.21	498	288,548			
Saskatoon/Regina CMA	1.30	856	324,433			
Saskatchewan Other	1.29	527	400,179			
Edmonton CMA	1.38	562	713,881			
Calgary CMA	1.31	553	700,911			
Alberta Other	1.39	435	762,709			
Vancouver CMA	1.41	632	1,538,851			
Victoria CMA	1.32	559	256,921			
British Columbia Other	1.35	362	1,250,085			

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 Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

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

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

Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of "travellers to British Columbia in the past two years" is more reliable than the estimated number of "travellers to British Columbia in the past two years. (Note that in the tables the CV's decline in value reading from left to right).

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

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

Rule 3: Estimates of Differences Between Aggregates or Percentages

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

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

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

Rule 4: Estimates of Ratios

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

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

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

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

Rule 5: Estimates of Differences of Ratios

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

The following 'real life' examples are included to assist users in applying the forementioned rules.

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

Suppose that a user estimates from the telephone survey data that 1,142,000 respondents from Quebec travelled to Ontario for a pleasure trip in the past two years. How does the user determine the coefficient of variation of this estimate?

1) Refer to the telephone survey CV table for Quebec. The estimated aggregate (1,142,000) 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 1,000,000.

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

3) So the approximate coefficient of variation of the estimate is 6.0%. The finding that there were 1,142,000 respondents from Quebec who travelled to Ontario for a pleasure trip in the past two years is publishable with no qualifications.

Example 2: Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates from the telephone survey data that 19.7% of female respondents from Halifax aged 60 years and older (4,042 out of 20,519) rent videos for a VCR on a regular basis. How does the user determine the coefficient of variation of this estimate?

1) Refer to the table for Halifax. Because the estimate is a percentage which is based on a subset of the total population (i.e., female respondents from Halifax aged 60 years and older), it is necessary to use both the percentage (19.7%) and the numerator portion of the percentage (4,042) in determining the coefficient of variation.

2) The numerator, 4,042, does not appear in the left-hand column (the 'Numerator of Percentage' column) so it is necessary to use the figure closet to it, namely 4,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 20.0%.

3) The figure at the intersection of the row and column used, namely 38.8% is the coefficient of variation to be used.

4) So the approximate coefficient of variation of the estimate is 38.8%. The finding that 19.7% of female respondents from Halifax aged 60 years and older rent videos for a VCR on a regular basis is considered unacceptable and Statistics Canada recommends this 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 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates from the telephone survey that 51.4% of male respondents from Edmonton (179,168 out of 348,702) said that they very likely or fairly likely to take a pleasure trip to the United States of one or more nights in the next two years, while the percentage was 37.8% for female respondents from Edmonton (138,114 out of 365,179). How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the Edmonton CV table in the same manner as described in Example 2 gives the CV of the estimate for males as 6.6%, and the CV of the estimate for females as 8.4%.

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

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

where $\hat{X}_1$ is estimate 1 (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 $\hat{d} = (.514-.378) = .136$ is:

$$\sigma_{\hat{d}} = \sqrt{[(.514)(.066)]^2 + [(.378)(.084)]^2}$$

$$= \sqrt{(.0011508) + (.0010081)}$$

$$= .046$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}}/\hat{d} = .046/.136 = 0.338$.

4) So the approximate coefficient of variation of the difference between the estimates is 33.8%. 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 from the telephone survey that 21,060 Winnipeg females aged 60 and older regularly read fashion or homemaking magazines such as Chatelaine, Vogue, Women's Day. An estimated 5,319 Winnipeg males 60 and older do the same. The user is interested in comparing the estimate of women versus men in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of Winnipeg females aged 60 and older who regularly read fashion or homemaking magazines and the denominator ($\hat{X}_2$) is the number of Winnipeg males with this same characteristic.

2) Refer to the table for Winnipeg.

3) The numerator of this ratio estimate is 21,060. The figure closest to it is 21,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 21.5%.

4) The denominator of this ratio estimate is 5,319. The figure closest to it is 5,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 45.0%.

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,

$$\alpha_{\hat{R}} = \sqrt{(.215)^2 + (.450)^2}$$

$$= \sqrt{(.046225) + (.2025)}$$

$$= 0.499$$

The obtained ratio of women versus men who regularly read fashion or homemaking magazines is 21,060 / 5,319 = 4.0. The coefficient of variation of this estimate is 49.9%, which should not be released under any circumstances. Rather, it should be deleted and replaced by dashes.

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 differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

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

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

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

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

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

Example of using the CV tables to obtain confidence limits:

A 95% confidence interval for the estimated proportion of respondents from British Columbia aged 18 and over who took an overnight business trip in the past two years would be calculated as follows.

$\hat{X}$ = 17.5% (or expressed as a proportion = .175)

t = 2

$\alpha_{\hat{X}}$ = 2.3% (.023 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

CI_X = {.175 - (2) (.175) (.023), .175 + (2) (.175) (.023)}

CI_X = {.175 - .008, .175 + .008}

CI_X = {.167, .183}

With 95% confidence it can be said that between 16.7% and 18.3% of respondents from British Columbia aged 18 and over took an overnight business trip in the past two years (using the reference year of the survey).

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 $\mathbf{X_1}$ and $\mathbf{X_2}$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

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

Example of using the CV tables to do a t-test:

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of male respondents from Edmonton who said that they very likely or fairly likely to take a pleasure trip to the United States of one or more nights in the next two years and the proportion of female respondents from Edmonton. From Example 3, Section 10.1, the standard error of the difference between these two estimates was found to be = .138. Hence ,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.514 - 0.378}{0.046} = \frac{0.136}{0.046} = 2.96.$$

Since t = 2.96 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for Travel Activities and Motivation Survey (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. For example, the coefficient of variation of the total number of trips to the Yukon in the past two years would be greater than the coefficient of variation of the corresponding proportion of persons who made a trip to the Yukon in the past two years. Hence if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

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

The minimum size of the estimate are specified in the table below by province for households. Estimates smaller than the minimum size given in the "Unacceptable" column must be flagged in the appropriate manner.

Table of Release Cut-offs (i.e. minimum size of an estimate) for the Telephone Survey

Area	Unacceptable CV > 33.3%	Marginal CV 16.6% to 33.3%	Acceptable CV 0.0% to 16.5%
Canada	**under 25,200**	**25,200 to < 102,300**	**102,300 and over**
Atlantic Provinces	under 24,400	24,400 to < 95,500	95,500 and over
Quebec	under 40,200	40,200 to < 160,300	160,300 and over
Ontario	under 21,100	21,100 to < 85,300	85,300 and over
Manitoba	under 8,000	8,000 to < 31,800	31,800 and over
Saskatchewan	under 6,600	6,600 to < 26,300	26,300 and over
Alberta	under 16,700	16,700 to < 66,500	66,500 and over
British Columbia	under 31,900	31,900 to < 125,700	125,700 and over
St. John's/Atlantic Other CMA	under 27,900	27,900 to < 107,600	107,600 and over
Halifax CMA	under 6,600	6,600 to < 25,100	25,100 and over
Montreal CMA	under 41,200	41,200 to < 160,100	160,100 and over
Quebec CMA	under 13,200	13,200 to < 50,400	50,400 and over
Quebec Other (excluding Hull)	under 51,600	51,600 to < 196,600	196,600 and over
Ottawa-Hull CMA	under 9,900	9,900 to < 38,800	38,800 and over
Toronto CMA	under 23,400	23,400 to< 93,300	93,300 and over
Hamilton CMA	under 5,700	5,700 to < 22,300	22,300 and over
London CMA	under 6,100	6,100 to < 23,400	23,400 and over
Kitchener CMA	under 5,200	5,200 to < 20,100	20,100 and over
St Catharines-Niagara CMA	under 4,000	4,000 to < 15,500	15,500 and over
Windsor CMA	under 4,300	4,300 to < 16,600	16,600 and over
Oshawa CMA	under 2,700	2,700 to < 10,800	10,800 and over
Sudbury/Thunder Bay CMA	under 2,600	2,600 to < 10,300	10,300 and over
Ontario Other	under 35,500	35,500 to < 138,500	138,500 and over
Winnipeg CMA	under 9,100	9,100 to < 35,200	35,200 and over
Manitoba Other	under 6,200	6,200 to < 23,600	23,600 and over
Saskatoon/Regina CMA	under 4,400	4,400 to < 17,100	17,100 and over
Saskatchewan Other	under 8,600	8,600 to < 33,000	33,000 and over
Edmonton CMA	under 15,500	15,500 to < 59,100	59,100 and over
Calgary CMA	under 14,700	14,700 to < 56,100	56,100 and over
Alberta Other	under 21,400	21,400 to < 80,100	80,100 and over
Vancouver CMA	under 30,400	30,400 to < 116,600	116,600 and over
Victoria CMA	under 5,400	5,400 to < 20,500	20,500 and over
British Columbia Other	under 40,700	40,700 to < 150,600	150,600 and over

Table of Release Cut-offs (i.e. minimum size of an estimate) for the Mail-out Survey

Area	Unacceptable CV > 33.3%	Marginal CV 16.6% to 33.3%	Acceptable CV 0.0% to 16.5%
Canada	**under 89,300**	**89,300 to < 358,700**	**358,700 and over**
Atlantic Provinces	under 71,100	71,100 to < 251,900	251,900 and over
Quebec	under 130,100	130,100 to < 487,100	487,100 and over
Ontario	under 88,300	88,300 to < 347,300	347,300 and over
Manitoba	under 27,600	27,600 to < 100,100	100,100 and over
Saskatchewan	under 20,400	20,400 to < 75,700	75,700 and over
Alberta	under 60,600	60,600 to < 225,500	225,500 and over
British Columbia	under 117,700	117,700 to < 424,200	424,200 and over

10.6 Coefficient of Variation Tables

Please refer to the files identified below for the coefficient of variation (CV) tables for the TAMS microdata:

TAMS1999_M_CVTabsE.pdf (Mail survey)

TAMS1999_T_CVTabsE.pdf (Telephone survey)

11.0 Weighting

This section outlines the derivation of survey weights for Travel Activities and Motivation Survey (TAMS). The weighting was done for the telephone survey and then the mail. The essential strategy of non-response adjustment remained the same throughout. The survey steps important for weighting were:

1) Generate a simple random sample of telephone numbers within 28 geographic strata;
2) Establish if the telephone number is a working residential telephone number;
3) Obtain a household roster;
4) Complete a telephone interview with a randomly selected person within household; and
5) Get a completed mail questionnaire for the travelers who completed the telephone survey.

To take into account the information we get at each step, we adjusted the initial probability of selection to arrive at the survey weights. For each step, we made the following adjustments:

1) Household non-response adjustment at the Regional Office level;
2) Adjustment for household size and number of telephones in household at the household level;
3) Person non-response adjustment at the household size level; and
4) Mail non-response adjustment using an age/sex/socioeconomic status/country of birth grouping.

Note that the Regional Office level was defined as Montreal, Sturgeon Falls, Winnipeg and Vancouver; and the household size level was one person household versus the two or more person household.

Initially at least, not much information was available for adjustments but as we moved from one step to the next, more and more data became available. By the time the mail non-response adjustment was made, we considered all of the telephone variables as potential candidates for making the mail adjustment. In the end, we defined the socioeconomic status as "lower," "middle," and "upper" based on reported education and income; the three groupings of age and sex were based on differential response patterns; and respondents born outside of the country were grouped together. In essence, we created response homogeneity groups to minimize measurable non-response bias.

After these adjustments were applied, we post-stratified to population totals. For the telephone survey, we post-stratified in a step-wise manner to (1) age-sex stratum totals; and (2) age-sex totals within province derived from 1996 Census totals from the Demographic Division of Statistics Canada. For the mail survey, we post-stratified to the estimated number of travelers per stratum based on stratum traveling rates. The end result was two survey weights, one for the telephone survey and one for the mail survey.

The strata are the 28 sub-provincial areas found in many of the tables in this document. We used the age groupings of 18-34, 35-49 and 50+ for the first post-stratification within stratum. For the age-sex breakdown within province we did not use province exactly but rather the grouping of Atlantic provinces, Quebec, Ontario, Manitoba, Saskatchewan, Alberta and British Columbia. The age breakdowns were 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69 and 70+.

Now that the process has been explained in words, let's move on to a more mathematical treatment of the weighting. We use the following notation:

Let	h	indicate the stratum, where h =1, 2, ..., 28;
	π_h	be the initial probability of selection of telephone numbers in stratum h;
	i	be the household with a completed roster;
	t_i	be the number of telephones in household i;
	n_i	be the number of eligible persons in household i;
	j_1	be the household adjustment class by regional office, j_1 =1, 2, 3, 4;
	$a\,j_1$	be the household non-response adjustment with subscript j_1;
	j_2	be the person adjustment class by household size, j_2 =1, 2;
	$a\,j_2$	be the person non-response adjustment with subscript j_2; and
	$I(\,)$	be an indicator variable equal to 1 if the condition is fulfilled, 0 otherwise;
	w^*	denote stepwise survey weights depending on the number of asterisks;

w^{tel} be the final telephone weight for completed telephone respondents; and
w^{mail} be the final mail weight for completed mail responses.

For the telephone two-step post-stratification, let

k_1 be the first telephone post-stratum, k_1 =1, 2, 3, ..., 174;
Mk_1 be the first telephone post-stratum total with subscript k_1 ;
k_2 be the second telephone post-stratum, k_2 =1, 2, 3, ..., 154; and
Mk_2 be the second telephone post-stratum total with subscript k_2.

Finally, for the mail sample let

l_1 be the mail adjustment class for the mail;
al_1 be the mail non-response adjustment with subscript l_1 ;
l_2 be the mail post-stratum; and
Ml_2 be the mail post-stratum total with subscript l_2.

Let's begin by defining the household non-response adjustment (the sums for the formulas below are over all in-scope telephone numbers), the initial person weight (takes into account the initial sampling probability, the number of telephones in a household and the number of eligible people in a household) and the person non-response adjustment:

$$a_{j_1} = \frac{\sum \pi_h^{-1} \; I\,(RO=j_1 \;\&\; in\text{-}scope\ telephone\ number)}{\sum \pi_h^{-1} \; I\,(RO=j_1 \;\&\; in\text{-}scope\ telephone\ number \;\&\; roster\ obtained)}$$

$$w_{hi} = \left(\frac{1}{\pi_h}\right)\left(\frac{n_i}{t_i}\right)$$

$$a_{j_2} = \frac{\sum_h \sum_i \sum_{j_1} w_{hi}\, a_{j_1} \; I\,(hhld\ size=j_2 \;\&\; roster)}{\sum_h \sum_i \sum_{j_1} w_{hi}\, a_{j_1} \; I\,(hhld\ size=j_2 \;\&\; roster \;\&\; complete\ telephone\ interview)}$$

The last two steps in obtaining the telephone weight is to post-stratify. The first of these steps adjusts the weights so that the weighted total of telephone responses adds up to the stratum control totals. We know, however, that the control totals, derived from demography estimates, at the sub-provincial level are not as accurate as the provincial age-sex control totals. This is why we first post-stratify to the stratum total and then use the provincial level age-sex control totals to do the final telephone post-stratification.

The post-stratification steps are simple adjustments by the ratio of the estimated control totals to the actual control totals. To simplify the notation, we've dropped all subscripts in the telephone weight except for subscripts *i* and *j*, referring to the responding household within stratum. The two post-stratification steps then look as follows:

$$\hat{M}_{k_1} = \sum_h \sum_i \sum_{j_1} \sum_{j_2} w_{hi}\, a_{j_1}\, a_{j_2}\; I(\text{stratum}=k_1)$$

$$\hat{M}_{k_2} = \sum_h \sum_i \sum_{j_1} \sum_{j_2} \sum_{k_1} w_{hi}\, a_{j_1}\, a_{j_2} \left(\frac{M_{k_1}}{\hat{M}_{k_1}}\right)$$

$$w_{hi}^{tel} = w_{hi}\, a_{j_1}\, a_{j_2} \left(\frac{M_{k_1}}{\hat{M}_{k_1}}\right) \left(\frac{M_{k_2}}{\hat{M}_{k_2}}\right) \qquad \forall\ i \in \text{responding telephone sample in stratum } h.$$

For the mail weight, we adjust for mail non-response using the response homogeneity groups described earlier and then, using derived stratum traveler control totals Ml_2, come up with the final weight. The sums are over all the telephone survey respondents:

$$a_{l_1} = \frac{\sum_h \sum_i w_{hi}^{tel}\; I(\text{socio-eco} * \text{agesex} * \text{birth country}=l_1\ \&\ \text{traveller})}{\sum_h \sum_i w_{hi}^{tel}\; I(\text{socio-eco} * \text{agesex} * \text{birth country}=l_1\ \&\ \text{traveller}\ \&\ \text{mail response})}$$

$$M_{l_2} = \sum_h \sum_i w_{hi}^{tel}\; I(\text{stratum}=l_2\ \&\ \text{telephone traveller})$$

$$\hat{M}_{l_2} = \sum_h \sum_i \sum_{l_1} w_{hi}^{tel}\, a_{l_1}\; I(\text{stratum}=l_2\ \&\ \text{mail response})$$

$$w_{hi}^{mail} = w_{hi}^{tel}\, a_{l_1} \left(\frac{M_{l_2}}{\hat{M}_{l_2}}\right) \qquad \forall\ i \in \text{responding mail-out sample in stratum } h.$$

Once again, the subscripts l_1 and l_2 have been dropped for the sake of clarity in the final step. Note that the definition of traveller for the above equations is a person who responded that they are a traveller to the telephone questionnaire (hence “telephone traveller”).

12.0 Questionnaires and Code Sheets

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

TAMS1999_M_QuestE.pdf (Mail questionnaire)

TAMS1999_T_QuestE.pdf (Telephone questionnaire)

13.0 Record Layout and Univariates

Please refer to the files identified below for the data files and codebooks for the Travel Activities and Motivation Survey (TAMS) microdata:

TAMS1999_M_CdBk.pdf (Mail codebook)

TAMS1999_T_Cdbk.pdf (Telephone codebook)