
AIR EXIT SURVEY

U.S. AND OVERSEAS VISITORS TO CANADA
ARRIVING AND DEPARTING BY COMMERCIAL AIR

PUBLIC USE MICRODATA FILE (PUMF) USER GUIDE

2024



How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to:

Statistics Canada

150 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6

Telephone: 1-800-263-1136

Email: infostats@statcan.gc.ca

Accessing and ordering information

The Air Exit Survey (AES) produces two types of microdata files - master files and public use microdata files (PUMF).

Master files

The master files contain all variables and all records from the survey collected during a reference period. These files are accessible at Statistics Canada for internal use and are also subject to custom tabulation requests.

Custom tabulations

Staff in the Client Services section of the **Tourism Statistics Program** can prepare custom tabulations based on the Master file. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Statistics Canada.

Public use microdata files (PUMF)

The public use microdata files are developed from the master files using a technique that balances the need to ensure respondent confidentiality with the need to produce the most useful data possible. The PUMF must meet stringent security and confidentiality standards required by the *Statistics Act* before they are released for public access. To ensure that these standards have been achieved, each PUMF goes through a formal review and approval process at Statistics Canada. Variables most likely to lead to identification of an individual are deleted from the data file or are collapsed to broader categories.

Note that there may be differences between the estimates published by Statistics Canada in its statistical products and the estimates reproduced by users with the public use microdata files, as Statistics Canada uses master files to produce its estimates. To obtain a copy of the PUMF, contact Statistics Canada.

The Data Liberation Initiative

The Data Liberation Initiative (DLI) Program enables students and researchers to use the public use microdata files in several universities across Canada. For more information, please consult the following web page: <http://www.statcan.gc.ca/eng/dli/dli>

Table of Contents

1.0 BACKGROUND	5
2.0 DATA COLLECTION	5
3.0 QUESTIONNAIRE CONCEPTS AND DEFINITIONS	5
3.1 Trips.....	6
3.2 Visits during a trip	7
3.3 Additional trip and visit characteristics.....	8
3.4 Travellers	9
3.5 Travel parties	10
4.0 DATA PROCESSING.....	11
4.1 Editing.....	11
4.2 Imputation	11
4.3 Disclosure control	12
5.0 WEIGHTING.....	12
6.0 ESTIMATION	13
7.0 SAMPLING VARIABILITY	13
7.1 Data quality.....	14
7.2 Bootstrap method of variance estimation	15
8.0 PUBLIC USE MICRODATA FILE (PUMF) CONTENT	15
Appendix A: List of standard geographical regions (SGR) for place of residence and their constitution for residents of overseas countries	17
Appendix B: U.S. state of residence groupings used in the 2024 Air Exit Survey PUMF	18
Appendix C: Bootstrap variance estimation for PUMF.....	19

1.0 BACKGROUND

The Visitor Travel Survey (VTS) is conducted by Statistics Canada to meet the requirements of the Balance of Payments (BOP) of the Canadian System of National Accounts. The survey is also used by the Tourism Satellite Account (TSA), Canada Border Services Agency (CBSA), Destination Canada, provincial and territorial tourism agencies and the private sector, and for reporting to international organizations such as United Nations Tourism and the Organization for Economic Co-operation and Development (OECD).

The VTS provides a full range of statistics on the volume of international travellers and detailed characteristics of their trips such as expenditures, activities, places visited and length of stay.

Prior to 2018, information about visitors to Canada was collected through the International Travel Survey (ITS). Through the years, questions directed at collecting detailed characteristics of travellers for market research and industry planning were gradually incorporated in the ITS. In 2018, the ITS was divided into the National Travel Survey (NTS), which covers international and domestic travel by Canadian residents, and the VTS, which covers travel to Canada by international (United States and overseas) residents.

The Air Exit Survey (AES) has been a component of the ITS or VTS since 2000, covering visitors departing by commercial air. Prior to the COVID-19 pandemic, data was also collected from non-air travellers via a distribution of invitation cards at border points. Public health restrictions forced all VTS collection activities to cease from March 2020 to March 2023, and while AES collection resumed in April 2023, data collection for travellers by modes other than commercial air did not, due to a low number of responses pre-pandemic.

Since a public use microdata file (PUMF) requires microdata from respondents, the 2024 iteration of the PUMF includes only the commercial air component collected by the AES and is therefore referred to as the 2024 AES PUMF. This manual has been produced to facilitate the use of this PUMF.

2.0 DATA COLLECTION

The AES collects monthly information on the characteristics of international travellers and trips, which are released quarterly. These details include place of residence, purpose of trip, travel party size, places visited, types of accommodation used, activities participated in during the trip, length of trip and trip spending.

AES collection targets U.S. and overseas resident visitors to Canada departing by commercial air from the international airports of Vancouver, Calgary, Toronto-Pearson, Montreal-Trudeau and Halifax. Statistics Canada interviewers use tablets to conduct face-to-face interviews with travellers in the departure areas at these five airports as they await their return flight.

AES interviews are conducted each month for a period of up to seven days. About 13,100 interviews (8,700 overseas and 4,400 U.S.) were completed from January to December 2024.

The targeted overseas countries are those from which Canada attracts the most visitors, including the United Kingdom, France, Germany, Mexico, Australia, China, Hong Kong, Japan, India, South Korea, the Netherlands and Switzerland. The AES team consists of interviewers fluent in various languages, enabling interviews to be conducted in the native language of the travellers where possible. A paper version of the questionnaire is also available in eight additional languages to facilitate tablet-based collection if needed.

3.0 QUESTIONNAIRE CONCEPTS AND DEFINITIONS

3.1 Trips

Trip

A trip begins when a non-resident traveller is cleared through a CBSA point of entry and enters Canada and ends when the traveller leaves Canada to return to their usual place of residence. Statistics Canada's Frontier Counts program uses data from CBSA to enumerate the number of trips made by international travellers each month, and data from this program are used to benchmark the trip information collected in the AES. A traveller may be included in the Frontier Counts more than once in the same reference period if they enter Canada more than once through a CBSA point of entry during this period.

Trip duration

The trip duration is measured by the number of nights the traveller spends in Canada. Trip durations are also divided into two mutually exclusive groups: same-day trips and overnight trips.

Same-day trip

Same-day travel encompasses travellers who enter and leave Canada on the same calendar day. This category includes persons residing in another country and commuting daily to and from work in Canada.

Overnight trip

Overnight travel encompasses travellers who stay in Canada for at least one night before exiting the country. This classification closely approximates the term "tourist" although it includes temporary and seasonal workers, students, and travellers who are in transit.

Purpose of trip

The main reason (or purpose) of a trip is categorized as follows in the 2024 AES PUMF:

- Holidays, leisure and recreation
- To visit friends or relatives
- To go to school or to study
- Other personal reason
- For a meeting, sales or service call
- Attend a conference, convention or trade show
- Other business reason
- In transit

In the past, and in some other AES or VTS products, these reasons have been aggregated as follows:

- Business (Includes attending a meeting, sales or service call, attending a conference, convention or trade show or seminar, and other work or business reasons).
- Visiting friends or relatives.

- Pleasure (Includes a holiday or vacation, visiting a second home, cottage or condo, and attending events and attractions).
- Other purposes (Includes going to school or studying, attending a non-business conference or convention, medical or health-related travel, shopping, transit and other personal reasons).

3.2 Visits during a trip

Visit

A non-resident traveller may visit several locations within Canada during the same trip. In the AES questionnaire, a respondent can declare up to ten places visited during their trip, and these are recorded at the Census Subdivision level. Note that for publishing purposes, a trip with multiple visits to the same geographical location will be counted as a single visit to that location.

Coding visits to geographical locations

Places visited (e.g., cities, towns, regions) are coded to the following standardized geographical classifications:

Census Metropolitan Area (CMA) and Census Agglomeration (CA)

CMAs are defined as the main labour market area of an urbanized core having 100,000 or more population. Smaller urban areas, with an urban core with populations of at least 10,000, are defined as CAs. For more details, refer to the Standard Geographical Classification SGC 2021.

Tourism Region (TR)

Tourism Regions are defined by the tourism departments of each province and territory as distinct areas of interest to the tourism industry. They are updated yearly.

Visit duration

AES respondents report the number of nights spent in each location visited. For publishing purposes, visit durations are divided into two mutually exclusive groups: same-day visits and overnight visits.

Same-day visit

A same-day visit is when the respondent has not declared any nights for a location visited. During a same-day trip, a traveller may visit several locations, however, only one location is captured on a same-day trip. For overnight trips, every location visited where the respondent did not spend a night is recorded as a same-day visit.

Overnight visit

An overnight visit is when the respondent has declared one or more nights in the visited location. Overnight visits can only be made while on an overnight trip.

3.3 Additional trip and visit characteristics

Mode of transportation

For the 2024 AES PUMF, all travellers entered and exited Canada by commercial air.

Route of entry

Non-resident travellers are grouped by the type of entry used to enter Canada. There are three categories: from the United States only, direct from another country, and from another country via the United States.

Accommodation type

The types of accommodation used in each location visited is collected in the questionnaire, and a respondent can report multiple types of accommodation for each visit.

The available categories are:

- Hotel
- Motel
- Home of friends or relatives
- Camp or trailer park
- Cottage or cabin
- Other type of accommodation

Activities

Respondents report all activities in which they have taken part during the trip according to provided categories. All categories are contained within two broader groupings: participating in sports or outdoor activities, or other activities.

No information is collected on activities during specific visits. For this reason, an activity cannot be associated with a precise location. For example, a person may visit Vancouver and Whistler during the same trip and declare they have done some downhill skiing. When compiling data, the “downhill skiing” activity will be attributed to both Vancouver and Whistler since activity information is only collected at the trip level. However, this activity may have occurred only in Vancouver or only in Whistler.

Expenditures

Trip expenditures made by all members of the travel party are collected in the questionnaire. Expenditures are not reported by visit, but rather by category for the entire trip. Expenditures may have been paid for by individuals, by government or by a private sector business, and are reported in Canadian dollars or are converted to Canadian dollars.

The expenditures are broken down into the following categories:

- Accommodation
- Transportation in Canada (including gasoline, car rental, intercity plane, bus and train fares, boat tours, local transit, taxi and rideshare)
- Food and beverage (including restaurants, bars and stores)

- Recreation and entertainment
- Clothing and gifts
- Other expenditures

Fares

Fares are the amounts spent by all members of the travel party on commercial flights to enter and exit Canada. Fares are reported in Canadian dollars or are converted to Canadian dollars. They are not part of expenditures, so are reported separately.

3.4 Travellers

International traveller

The term international traveller applies to all persons arriving in Canada who are cleared through CBSA points of entry. The Frontier Counts enumerates each entry into Canada, including entries by non-residents and re-entries by residents. Entries are enumerated rather than unique travellers, so a traveller will be counted each time they cross the border during the reference period.

The term international traveller is divided into three mutually exclusive groups: non-resident traveller, resident traveller and other traveller. The AES PUMF only includes non-resident travellers. For documentation on resident travellers, refer to the National Travel Survey (NTS) Microdata User Guide.

Non-resident traveller

The AES defines a non-resident traveller as an international traveller whose country of residence is a country other than Canada and who is cleared through a CBSA point of entry on a visit for a period of less than eight months. Canadian citizens who have resided in another country for more than 12 months and are visiting Canada are considered non-resident travellers.

Resident traveller

A resident traveller is any Canadian resident (both Canadian and foreign citizens) who has travelled outside of Canada for a period of less than 12 months and who is returning to Canada through a CBSA point of entry. These travellers are the target population for the NTS and are not part of the AES PUMF.

Other traveller

Other travellers are excluded from the target population of the AES since they are not travelling for a tourism-related purpose, and include the following types of international travellers:

- **Immigrants:** Citizens of other countries entering Canada to take up permanent residence for more than 12 months.
- **Former residents:** Canadian citizens returning to Canada to re-establish permanent residence after residing outside Canada for more than 12 months.

- **Military personnel, diplomats and dependents:** Canadian citizens and dependents who have been stationed abroad for more than 12 months and who are returning from postings. Foreign citizens and dependents entering Canada on military or diplomatic postings.
- **Crew:** Persons engaged in the operation of a commercial aircraft.

3.5 Travel parties

A travel party refers to persons from the same country who are travelling together (whether they are from the same household or not) and who are able to collectively report their trip spending and activities.

Travel party composition

The travel party composition refers to the number of adults and/or children who are part of the same travel party.

Place of residence (Place of origin)

The place of residence (or place of origin) is the place of last permanent residence. In the 2024 AES PUMF, for residents of the United States of America this is the state (or state grouping) of residence, and for overseas residents it is the country (or country grouping).

Person-trip

A person-trip begins when a non-resident traveller is cleared through a CBSA point of entry and enters Canada and ends when the traveller leaves Canada to return to their usual place of residence. A person-trip differs from a trip in that a person-trip is weighted by the travel party size, while a trip is taken by the whole travel party.

Person-trip-night

A person-trip-night is defined as a night a non-resident traveller spends within Canada during a person-trip.

Person-visit

A person-visit is either a same-day or overnight visit made by an individual taking a trip. A person-visit differs from a visit in that a person-visit is weighted by the travel party size, while a visit is taken by the whole travel party. The total number of person-visits is the total number of visits made by each person, in each travel party, on each trip taken.

Person-visit-night

A person-visit-night is a night spent in the visited location by a person taking a trip. For example, a person who spends three nights in the visited location has a count of three person-visit-nights in that location, as do three people who spend one night in the visited location.

4.0 DATA PROCESSING

Survey processing occurs after all data has been collected and prior to estimation, and involves data editing, coding, imputation and verification. It serves to make the data suitable for estimation and tabulation.

4.1 Editing

Data is not always reported on a questionnaire in such a way that it can be used directly in further analysis. Editing reformats data so that it can be used in tabulation, analysis, and estimation.

4.1.1 Validity and consistency edits

Some errors occur in the data capture process such as typos or missing data on the questionnaire. These errors must be located and corrected. Edit rules are designed to ensure the data is consistent with known characteristics, and that questionnaire entries are valid responses given the question.

Some examples of edits made during the data capture process follow:

- All respondents to the AES are leaving Canada by commercial air, so even if a respondent indicates leaving by car, this is corrected to be commercial air.
- If all transportation costs are included in a package cost but are also included in the entries for fare to enter Canada and fare to leave Canada, then the entries are corrected to remove the duplication of reported fares.

4.1.2 Derived variable creation

Expenses reported on package tours must be broken down into spending components, reported data for trips that involved multiple entries into Canada must be separated, and derived variables must be generated.

The AES asks respondents for information on packages and inclusive trips purchased. These purchases are not reported in the expenditure within Canada section. Consequently, packages are separated according to the components covered within the collective package cost and added to the relevant expenditure categories. The expenditure variables that include the package and non-package values are referred to as global expenditure variables.

From the AES definition of a trip, any exit from Canada marks the end of a trip. Respondents might report several trips under this definition within their completed questionnaire. These trips must be separated to allow for consistent analysis.

Finally, all variables that are required in further analysis are created.

4.2 Imputation

Data imputation is performed in the AES whenever fields are missing within a questionnaire, or when an insufficient number of questionnaires are received to meet the AES objectives.

A minimum amount of information is required for a questionnaire to be considered reliable. The following information must be available for a questionnaire to be retained and to contribute to the survey estimates:

- Port of entry
- Mode of entry
- Route of entry
- Date of entry
- Date of exit
- Purpose of trip
- Travel party size
- Places visited
- Whether a package was purchased and the corresponding nights included in the package

These completed fields enable a questionnaire to be linked to a calibration group¹ to impute missing values. For example, missing transportation fares and/or total trip spending are imputed using regression models based on the travel party size, trip duration, purpose of trip, ages of travel party members, accommodation information and port of entry.

Total imputation (i.e. imputation of entire questionnaires) is performed for any calibration group for which an insufficient number of questionnaires was completed for a given quarter. Since estimates are generated by calibration group, a minimum number of questionnaires is required to provide quality estimates for the calibration group. In instances where the minimum number is not met by received questionnaires from the current quarter, questionnaires are added (imputed) to the calibration group for the reference quarter. These imputed questionnaires are selected from questionnaires from the same quarter in previous survey years. For the 2024 AES, the rates of total imputation are low, at less than 0.1% of records.

4.3 Disclosure control

All survey respondents are assured that their identities and responses will remain confidential and that no identifying information will be disclosed in the PUMF. This is a requirement by law to which all Statistics Canada data adheres. Any information that may divulge a respondent's identity, without prior knowledge or written consent from the respondent, remains confidential and cannot be released or published.

Some data are therefore suppressed, perturbed or removed entirely from the PUMF to prevent any direct or residual identification of respondents from the released data. Variables or categories of variables that can link a response to an individual are removed or suppressed. This includes variables with categories selected by a small number of respondents or providing identifiable information such as visits made during a trip, which can be traced back to a respondent.

Programs are run to ensure the confidentiality of all released data following the disclosure control measures that are applied to the PUMF.

5.0 WEIGHTING

For the U.S. traveller flow, the final weights are calibrated to match the totals reported by Statistics Canada's Frontier Counts program, which is an administrative data program based on border crossing

¹ In the Air Exit Survey, the target populations (international travellers) are partitioned into calibration groups, based on selected traveller characteristics, such as country of residence and duration of stay.

information from CBSA. The final calibration groups are defined by five Canadian provinces or regions of entry (Atlantic, Quebec, Ontario, Prairies, and British Columbia/North). Then, for Atlantic and Prairies, two trip duration categories (same-day and overnight) are defined, while for other groups, four trip duration categories are defined (same-day and three durations of overnight), resulting in 16 total calibration groups.

For the overseas traveller flow, each overseas traveller is coded by country of residence (or, in some instances, country groupings) to obtain a standard geographical region (SGR) code. Some of these SGR codes are further collapsed to form grouped SGRs, of which there are 24, that are used as part of the control groups for the 2024 AES. Duration (short and long) counts, the other part of the control groups, are estimated based on Primary Inspection Kiosk information from CBSA and Frontier Counts. For different SGRs, the short and long definitions are different. The initial weights for the overseas traveller flow are calculated for each grouped SGR by duration, resulting in 48 calibration groups in total. The final weights of the overseas travellers are calibrated to match five Canadian regions by duration totals (10) and grouped SGR totals (24), resulting in 34 final calibration groups. Note that some of the final calibration groups overlap.

6.0 ESTIMATION

For estimation purposes, the responses obtained through the questionnaires must be treated as a simple random sample from the total traffic in each calibration group. The data are subject to some degree of “distribution bias” since the distribution of respondents may be different from the distribution of the population. For example, many travellers with a particular spending habit may respond to the survey, but if that spending habit isn’t very common in the population, then there’s distribution bias. The data are also subject to some “non-response bias” since the individuals who choose not to reply may share a characteristic. This may make the responses to some degree unrepresentative of the travelling population.

Weighting techniques used in the estimation process attempt to reduce the effect of biases. Questionnaires are reorganised by known characteristics into homogeneous groups. Using the calculated weights, estimates can be obtained for a variety of trip and traveller characteristics. These estimates take the form of totals, averages and percentages, and can be obtained if a sufficient record sample exists for the desired grouping of traveller characteristics and geography.

7.0 SAMPLING VARIABILITY

The AES is based on data gathered from only a part (sample) of the population. It is important to mention that results are only estimates of “true” values, where the “true” values for the travelling population could only be obtained through a census of all travellers. Since differences always exist between sample and census data, statistical measures have been developed which indicate the expected size of the differences. Such measures allow analysts to assess the reliability of statistics produced from sample surveys.

Various factors contribute to the quality of an estimate, such as distribution bias or non-response bias. The size of the sample being used to produce an estimate is important for quality estimates; if the number of respondents with the characteristics of interest is too small, particularly relative to the census volume, then the estimate will be less reliable.

7.1 Data quality

Data quality is affected by both sampling and non-sampling errors. For estimates derived from the AES PUMF, the quality level of an estimate will be determined only based on sampling error, since it is difficult to measure non-sampling error. Before making inferences or publishing any estimates from the AES PUMF, users should first determine the quality level of the estimates.

First, the number of records contributing to the estimate should be determined. In general, if this number is less than 10, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 10 or more, users should calculate a measure of variability, such as the Coefficient of Variation (CV) of the estimate and release it alongside the estimate itself. Release guidelines recommended by Statistics Canada are provided in the next section and should be applied to rounded weighted estimates.

Most estimates can be considered releasable. However, those of marginal or unacceptable quality level should be accompanied by a warning to caution subsequent users. Estimates of low or unacceptable quality should, ideally, not be released, but if they are, they must be accompanied by a disclaimer.

7.2 Coefficient of variation and release guidelines

The recommended quality levels based on the CV of the estimate are Excellent (A), Very good (B), Good (C), Acceptable (D) and Use with caution (E).

Category A: Excellent

If the CV of an estimate falls between 0.0% and 5.0%, it is considered to be of excellent quality. No warnings or restrictions will be associated with the estimate in the data.

Category B: Very good

If the CV of an estimate falls between 5.01% and 15.0%, it is considered to be of very good quality. No warnings or restrictions will be associated with the estimate in the data.

Category C: Good

If the CV of an estimate falls between 15.01% and 25.0%, it is considered to be of good quality. The estimate is released as it may be useful for some purposes but is displayed with a warning regarding the accuracy of the estimate.

Category D: Acceptable

If the CV of an estimate falls between 25.01% and 35.0%, it is considered to be of acceptable quality. The estimate is released as it may be useful for some purposes but is displayed with a warning regarding the accuracy of the estimate.

Category E: Use with caution

If the CV of an estimate is 35.01% or higher, it is considered to be of unacceptable quality. The estimate should not be released as it does not meet the quality standards of Statistics Canada. Any conclusions made using an estimate of unacceptable quality is unreliable and likely invalid.

However, if a user insists on the inclusion of estimates considered to be of unacceptable quality, the estimates must be displayed with a disclaimer. The user is also required to always include the disclaimer if these estimates are included (either directly or indirectly) in data that is disseminated, reported, or presented. Any estimate calculated from fewer than 10 records is also considered to be of unacceptable quality, regardless of its corresponding CV.

7.2 Bootstrap method of variance estimation

Because of the complex nature of the AES design and processing, simple variance calculations using standard methods tend to under-estimate the variance, making the estimates appear more precise than they are. For the AES, Statistics Canada uses the bootstrap method of variance estimation. Under this method, the respondents are sub-sampled and re-weighted many times over (the AES currently uses 500 bootstrap sub-samples). From these sub-samples, an estimated variance is obtained to derive the CV.

Bootstrap weight files are provided with the PUMF in a CSV format, so that users can calculate variability measures for the estimates that they derive from the PUMF. Appendix C provides an overview of how to use bootstrap weights to do variance estimation.

8.0 PUBLIC USE MICRODATA FILE (PUMF) CONTENT

The AES PUMF contains information on the travel habits of visitors (U.S. and overseas residents) to Canada who entered and exited by commercial air. The following characteristics are available on the AES PUMF:

Trip characteristics:

- Quarter of entry
- Province or region of entry
- Mode of transportation used to enter and exit Canada, and modes used while in Canada
- Main purpose of trip
- Activities
- International transportation costs (fares paid to all carriers)
- Total trip spending (excluding fares)
- Spending by category
- Number of nights spent in Canada
- Number of nights by province
- Number of person-visits by province
- Place of residence (U.S. state or state grouping, and country or country grouping)
- Travel party size

- Weight of records

Demographic characteristics:

- Composition of travel party by age and sex groupings

Places visited:

- Place visited (province, Census Metropolitan Area, Tourism Region) (max. of 10)
- Length of stay in place visited (number of nights)
- Accommodations used in place visited

A complete listing of the variables on the PUMF and their frequencies can be found in the code book.

Note that as part of the disclosure control measures discussed in Section 4.3, the order of visits of each trip is usually adjusted, so users should consider the order of visits 1-10 in the file as having little analytical value. Users should focus on overall estimates for visit locations irrespective of their order in the trip, for which significant effort is made in the PUMF creation process to maintain analytical utility.

Appendix A: List of standard geographical regions (SGR) for place of residence and their constitution for residents of overseas countries

Refer to the file AES_PUMF_2024_UserGuide_Appendix.pdf for a list of countries included in each standard geographical region (SGR).

Appendix B: U.S. state of residence groupings used in the 2024 Air Exit Survey PUMF

Refer to the file AES_PUMF_2024_UserGuide_Appendix.pdf for the U.S. state of residence groupings that were used for the 2024 AES PUMF, including a list of states that make up the “other” regional categories, where applicable.

Appendix C: Bootstrap variance estimation for PUMF

To determine the quality of the estimates and calculate the coefficient of variation (CV), the standard error must be calculated. Confidence intervals also require the standard error of the estimate. The AES uses an atypical survey design and calibration, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, the bootstrap method for variance estimation, a pseudo-replication technique, was used to measure the quality of the AES estimates. With the use of the bootstrap weights and one of the available statistical packages for variance estimation, CVs and other variance estimates can be derived with accuracy.

Bootstrap weights are provided with the PUMF in a CSV format.

Bootstrap method for variance estimation

The mean bootstrap method was used for AES and for its PUMF. For each sample unit, a set of 500 final bootstrap weights are generated using replication methods, with each of these bootstrap weights being the result of averaging 20 bootstrap sub-replicates.

The variation among the 500 possible estimates based on each of the replicates is related to the variance of the estimator based on the survey weights and can be used to approximate it.

Users should note that the variance estimates from the PUMF bootstraps may be slightly different than the variance estimates released by Statistics Canada for the AES, as some small adjustments were made to protect respondent confidentiality.

Statistical packages for variance estimation

There are various commercial or open-source software packages that allow users to employ bootstrap replicate weights for variance estimation.

Using the bootstrap weights

Example: Estimating expenditures

Suppose that a user wants to estimate the amount that residents of the United Kingdom departing by commercial air spent on travel to Canada in 2024 (β).

The user may estimate β by multiplying the total expenditures (VGLTOTSP) for each trip by United Kingdom residents (VSGRCDEP = 101) by the corresponding trip weight (VWEIGHTP) and then summing these weighted expenditures. This gives the following result:

$$\hat{\beta} = \$1,435,576,664.$$

This process can be repeated in the same manner, except by substituting the trip weight (VWEIGHTP) by the b^{th} bootstrap weight, where $b=1, \dots, 500$ (BSW1, ..., BSW500).

This produces the bootstrap estimates $\hat{\beta}_{(1)}, \dots, \hat{\beta}_{(500)}$ of β which have been provided in the following table:

Bootstrap Weight	Bootstrap Estimate	Weighted Sum of Expenditures
BSW1	$\hat{\beta}_{(1)}$	\$1,442,000,001
BSW2	$\hat{\beta}_{(2)}$	\$1,445,986,600
BSW3	$\hat{\beta}_{(3)}$	\$1,435,088,937
BSW4	$\hat{\beta}_{(4)}$	\$1,433,135,899
BSW5	$\hat{\beta}_{(5)}$	\$1,441,369,465
...		...
BSW500	$\hat{\beta}_{(500)}$	\$1,434,615,316
Bootstrap Mean	$\hat{\beta}_B := \frac{1}{500} \sum_{b=1}^{500} \hat{\beta}_{(b)}$	\$1,434,546,374

Then the usual bootstrap estimate of the variance of $\hat{\beta}$ can be calculated as follows. Note that a factor of 20 must be applied to accurately calculate the variance, as each set of bootstrap weights (from BSW1 to BSW500) results from averaging 20 bootstrap sub-replicates using the mean bootstrap method:

$$\hat{V}_B(\hat{\beta}) = 20 \times \frac{1}{500-1} \sum_{b=1}^{500} (\hat{\beta}_{(b)} - \hat{\beta}_B)^2 = 2.1736311 \times 10^{15}$$

From this, one can estimate the standard error as:

$$\widehat{SE}_B(\hat{\beta}) = \sqrt{\hat{V}_B(\hat{\beta})} = 46,622,217$$

and the coefficient of variation as $CV_B(\hat{\beta}) = \frac{\widehat{SE}_B(\hat{\beta})}{\hat{\beta}} \times 100\% = 3.248\%$.