PUBLISHING GUIDELINES AND TIPS FOR USING TRA DATA

INTERNATIONAL VISITOR SURVEY, NATIONAL VISITOR SURVEY AND REGIONAL EXPENDITURE (REX) DATA

FREQUENTLY ASKED QUESTIONS (FAQS)

CONTENTS

INTRODUCTION

This document addresses frequently asked questions regarding International Visitor Survey (IVS) and National Visitor Survey (NVS) results published by Tourism Research Australia (TRA), including the calculations required to ensure consistency. The key methodological differences between the surveys and modelled data for regional expenditure (REX) are also discussed.

INTERNATIONAL

The IVS is a comprehensive collection of topics on international visitation to Australia by travellers aged 15 years and over.

EXPENDITURE

IVS expenditure is collected at the national (trip) level as the survey only asks the respondent details on expenditure for their entire trip to Australia. This data is available in CD-MOTA and TRA Online.

State and regional expenditure data should not be obtained through CD-MOTA via the IVS and NVS results. Regional expenditure estimates are derived using an expenditure model known as REX. However, REX results can be provided by TRA if expenditure information at the state or regional level is required.

PACKAGE EXPENDITURE

A series of studies by the ABS (1995, 1996), Australian Tourism Export Council (2000), Bureau of Tourism Research (Bonnet et al. 1994 and Skene, 1995) and Office of Economic and Statistical Research (2001) estimated the proportion of travel packages spent on items other than long distance fares to be between 26% and 35%. Based on these results, TRA distributes total package expenditure using a 30/70 split; 30% is attributed to expenditure in Australia while the remaining 70% is assumed to be for the purpose of long distance fares and thus spent in the visitor's home country.

IVS FAQS

IN REX, WHAT ITEMS ARE INCLUDED OR EXCLUDED FROM THE EXPENDITURE ESTIMATES?

The REX model for IVS estimates **excludes** the following expenditure items:

* international airfares, motor vehicles and other major equipment

* 70% of pre-paid package expenditure (as outlined in the package expenditure section above).

CAN I CROSS-CHECK EXPENDITURE IN CD-MOTA?

Yes, but **only** for the **Australia totals**. To complete a CD-MOTA cross check, create a derived item summing across all expenditure items and exclude international airfares, motor vehicles and other major equipment. Ensure you multiply package expenditure by 0.3 (removing the 70% prepaid package) as per the REX model.

WHY ARE PUBLISHED REX ESTIMATES FOR NIGHTS DIFFERENT TO THE CD-MOTA'S VISITOR NIGHTS NUMBER?

REX estimates exclude nights where a respondent was transiting and did not have a stopover.

CAN I CROSS-CHECK NIGHTS IN CD-MOTA?

Yes, but **only** for the **Australia totals**. To cross-check nights in CD-MOTA, include the summation option *Visitor Nights* and then subtract the value for *Weighted nights spent in transit* found under '*Other weighted values'*. The result will match the REX table value.

DO REX VISITOR ESTIMATES MATCH VISITOR NUMBERS IN CD-MOTA?

Visitor estimates data for Australia will match. However, state and region data may have discrepancies for areas containing arrival or departure ports. In these cases, the REX estimate can be slightly higher than the CD-MOTA value due to the inclusion of transit-only cases in the REX visitor count.

HOW DO I CALCULATE AVERAGE TOTAL TRIP EXPENDITURE ESTIMATES?

Simply divide the summation option *Itemised expenditure including pre-paid package and international airfare* by the summation option *Visitors*.

This creates an estimate for the average total trip spend for international travellers and includes some expenditure that remains overseas via the pre-paid component.

Note: The result will be in thousands—multiply by 1,000 to gain a whole dollar amount.

HOW DO I CALCULATE AVERAGE ESTIMATES FOR SPEND IN AUSTRALIA ONLY?

Simply divide the summation option *Itemised expenditure excluding international airfare and pre-paid package* by the summation option *Visitors*.

This creates an estimate for the average spend in Australia by international travellers and excludes all expenditure that remains overseas through the pre-paid components of international airfares and packages. It also excludes approximately 30% of expenditure that would have come to Australia under pre-paid packages.

Note: The result will be in thousands of dollars—multiply by 1,000 to obtain a whole dollar figure.

HOW DO I CALCULATE AVERAGE EXPENDITURE ESTIMATES PER VISITOR FOR A STATE, TERRITORY OR TOURISM REGION?

Using REX results, divide *Expenditure including package ($million)* by *Visitors (000's) (*ensure you use REX results for both). Multiply the result by 1,000 to scale to REX estimates*.*

This calculation creates an estimate for the average amount of money spent in any particular state, territory or tourism region in Australia by each international visitor.

Note: Due to the use of the variable *Expenditure including package ($million),* some allocated expenditure will be pre-paid spend that has not been spent in that state, territory or tourism region. The REX variable *Expenditure excluding package ($ million)* excludes these pre-paid monies and may be more relevant to the destination depending on the purpose of the research.

HOW DO I CALCULATE AVERAGE EXPENDITURE PER NIGHT ESTIMATES FOR A STATE, TERRITORY OR TOURISM REGION?

Using REX results, use *Expenditure including package ($million)* and divide by *Nights (000's)* (ensure you use the REX results for both). Multiply the result by 1,000 to scale to whole dollar amounts*.*

Note: Due to the use of the variable *Expenditure including package ($million),* some allocated expenditure will be pre-paid spend that has not been spent in that state, territory or tourism region. The REX variable *Expenditure excluding package ($million)* excludes these pre-paid monies and may be more relevant to the destination depending on the purpose of the research.

DOMESTIC

The National Visitor Survey (NVS) is a comprehensive collection of topics on travel by Australians aged 15 years and over.

EXPENDITURE

NVS expenditure is collected at the national (trip) level as the survey only asks the respondent expenditure details about the entire trip in Australia. This data is available in CD-MOTA and TRA Online.

State and regional expenditure data is not available in CD-MOTA or on TRA Online as it is derived using an expenditure model known as REX. However, REX results can be provided by TRA for users requiring expenditure information at the state or regional level.

NVS FAQS

IN NVS OVERNIGHT REX ESTIMATES, WHAT ITEMS ARE INCLUDED OR EXCLUDED FROM THE EXPENDITURE?

The REX model for domestic overnight trips produces estimates that exclude the following expenditure items:

- airfares, motor vehicles and other long distance transport

- 60% of pre-paid package expenditure (due to industry assumptions).

DO NVS OVERNIGHT ESTIMATES IN REX AND CD-MOTA MATCH?

That depends on the summation option used.

- CD-MOTA's *Expenditure ($000-excluding motor vehicles)* will match the REX *Expenditure – including airfares and long distance transport costs* estimates.

- CD-MOTA cannot match the REX variable *Expenditure –excluding airfares and long distance transport* exactly. Approximate data from CD-MOTA can be obtained by excluding airfares and transport costs and including 40% of the package expenditure (multiply package spend by 0.4).

- CD-MOTA results for visitors and nights should match REX results.

WILL REX DAY-TRIP ESTIMATES FOR EXPENDITURE AND TRIPS MATCH THE RESULTS IN CD-MOTA?

Published REX estimates for day trips and expenditure should match CD-MOTA.

HOW DO I CALCULATE AN AVERAGE EXPENDITURE PER TRIP ESTIMATE?

Using CD-MOTA for NVS overnight take *Expenditure – excluding motor vehicles ($000s)* and divide by the number of *Overnight trips ('000)*. For day trips use *Expenditure – excluding motor vehicles ($000s)* and divide by the number of *Daytrips ('000)*.

HOW DO I CALCULATE AN AVERAGE EXPENDITURE ESTIMATE FOR A STATE, TERRITORY OR TOURISM REGION ON OVERNIGHT VISITS?

Using REX results, take *the item Expenditure excluding airfares and other long distance transport costs ($million)* and divide by the REX result for *Visitors (000's)*.

Note: The result will be in thousands of dollars—multiply by 1,000 to show a whole dollar amount.

Using REX results, take *Expenditure excluding airfares and other long distance transport costs ($million)* and divide by *Visitors (000's)* (ensure you use the REX results for both). Multiply the result by 1,000 to scale to whole dollar amounts.

Depending on the research topic, *Expenditure including airfares and other long distance transport costs ($million)* can be used in the calculation. *Expenditure excluding airfares and other long distance transport costs* is generally more relevant to spend in the specific destination of that visit, as expenditure on airfares and other long distance transport are items that are typically paid for at the origin of the trip.

For NVS day trips from REX results, use '*Expenditure – excluding motor vehicles (000s)'* and divide by the total number of *Daytrips ('000)*.

HOW DO I CALCULATE AN AVERAGE EXPENDITURE PER NIGHT ESTIMATE FOR A STATE, TERRITORY OR TOURISM REGION?

Using REX results, take *Expenditure excluding airfares and other long distance transport costs ($million)'* and divide by *Visitor nights (000's)*.

Note: The result will be in thousands of dollars—multiply by 1,000 to show a whole dollar amount.

Depending on the research topic, *Expenditure including airfares and other long distance transport costs ($million)* can be used in the calculation. *Expenditure excluding airfares and other long distance transport costs* is generally more relevant to expenditure in the specific destination, as expenditure on airfares and other long distance transport are items that are typically paid for at the origin of the trip.

PUBLISHING IVS, NVS AND REX ESTIMATES

INTERNATIONAL & NATIONAL VISITOR SURVEYS

The National Surveys are sample-based surveys rather than a census. Therefore, results are subject to variability and sample error. As a guideline and to assist users, TRA publishes error rates expressed as a percentage calculated at the 95% confidence interval (see Attachment A).

While statistical disciplines/rules provide a guide, there is no substitute for having a good knowledge of the subject matter and an understanding of the survey principles.

For example, a cell with a sample size of 30 for the United Kingdom (UK) will most likely have a greater error than that of a cell with a sample of 30 for Japan. In fact, the cell sample result for Japanese tourists is most likely to be reasonable. This is due to the fact that we know Japanese travellers in Australia are a very homogenous group (following similar travel patterns and behaviours), whereas the UK traveller is more independent and disperses widely. Therefore, a larger sample is needed to get an acceptable result for the population.

TRA utilises guidelines to draw a line between estimates that are reliable and those that are most likely to have a sampling variability too high for most practical purposes. As shown in the example above, the guidelines are just that—there is no substitute for knowledge of the survey dataset. However, calculating the sample error for every individual estimate is not practical.

Survey data that falls below the standard guidelines for the IVS and NVS are generally not published by TRA. However, where possible, TRA can provide the data to users with the provision that caution is used. Affected cells should be highlighted via formatting (coloured text or flagging with symbols) and footnotes supplied as appropriate.

The standard TRA footnote is:

*Please use data with caution, this estimate has a sampling variability too high for most practical purposes.

GUIDELINES FOR PUBLISHING INTERNATIONAL VISITOR SURVEY ESTIMATES

HOW DO I DECIDE IF AN ESTIMATE SHOULD BE FLAGGED DUE TO HIGH SAMPLING VARIABILITY?

Any cell where the estimate has a sample size less than 40 and/or a confidence interval greater than 50% is considered to have high sampling variability.

In the absence of sample and/or confidence interval indicators, flag cells with high sampling variability using the table below.

For	using the estimate for...	flag cells less than…
IVS – Visitors	Visitor	5,000
IVS – Nights	Nights	250,000
IVS – Expenditure	Expenditure	$25 million
NVS – Overnight visitors	Visitor	50,000
NVS – Nights	Nights	500,000
NVS – Overnight expenditure	Expenditure	$50 million
NVS – Day trips	Trips	100,000
NVS – Day trip expenditure	Expenditure	$30 million
NVS – Outbound	Visitor	20,000
NVS – Outbound nights	Nights	250,000
NVS – Outbound expenditure	Visitor	$150 million

GUIDELINES FOR PUBLISHING REGIONAL EXPENDITURE (REX) ESTIMATES

REX is a model that is based on outputs from the national surveys (sample surveys). Variability within the REX estimates will be greater than that of the national surveys; therefore, the rules and restrictions around the publishing of REX estimates are much tighter. Estimates that fall under the guidelines are simply labelled 'np' (not publishable).

WHAT ARE THE PARAMETERS FOR PUBLISHING OR NOT PUBLISHING REX ESTIMATES?

Any cell where the estimate has a sample size less than 40 and/or a confidence interval greater than 50% should be populated with 'np' for not publishable.

ARE THERE ANY GUIDELINES FOR REX ON COMPARING TWO NUMBERS?

Common sense should be used when assessing the degree of change between two numbers. A movement that looks out of step with the trend estimates should be treated with caution. Contact TRA for further analysis of the data if the use of the number/estimate is essential.

For further information on the iterative process in regional expenditure estimation, please see explanatory notes at Attachment C.

ATTACHMENT A

IVS DATA RELIABILITY

The results given in the IVS are based on a sample, rather than a census, of international visitors to Australia. As with all sample surveys, the results are subject to sampling variability, and therefore differ from figures obtained if all international visitors to Australia had been surveyed.

A measure of the possible degree of difference is given by the relative standard error of the survey and its associated confidence interval, which indicates the extent to which an estimate might vary from the true figure.

The table below provides the 95% confidence interval widths for a range of estimates available in the IVS. That is, there are approximately 19 chances in 20 that the true number is within the range identified by applying the figures in the table.

Size of 95% confidence interval for estimate (expressed as a percentage of the estimate)

Estimate	Visits	Nights	Expenditure
2 000	64.4%	#	#
5 000	41.5%	#	#
10 000	29.8%	#	#
20 000	21.4%	#	#
50 000	13.8%	#	#
100 000	9.9%	75.7%	#
200 000	7.1%	55.4%	#
500 000	4.6%	36.7%	#
1 000 000	3.3%	26.9%	#
2 000 000	2.4%	19.7%	#
5 000 000	1.5%	13.1%	97.6%
10 000 000	1.1%	9.6%	72.8%
20 000 000		7.0%	54.3%
50 000 000		4.6%	36.8%
100 000 000		3.4%	27.4%
200 000 000		2.5%	20.5%
500 000 000		1.7%	13.9%
1 000 000 000		1.2%	10.3%
2 000 000 000		0.9%	7.7%
5 000 000 000		0.6%	5.2%
10 000 000 000		0.4%	3.9%

- 95% confidence interval is greater than estimate.

The following example illustrates the use of this table to determine a range within which we are 95% confident that the true total lies. Say, the estimated number of Chinese visitors who stayed in Queensland was 100,000. Looking at the visits column (see table), an estimate of 100,000 visitors has a 95% confidence interval of 9.9%. Therefore, we are 95% confident that the true number of Chinese who stayed in Queensland was between 90,100 and 109,900 visitors (100,000 ± 9.9%).

The IVS relative standard errors were calculated using the Complex Survey Sampling module in SPSS V14.0. Estimates of variation are based on sampling with replacement principles and make allowances for the IVS stratification. The covariance and estimates output from this program were then regressed with a log transformation using Ordinary Least Squares (OLS) regression to achieve three independent models (for visitors, nights and expenditure). The models were computed using the R statistical program and the actual relationship modelled was:

$1n (COV) = a + b * 1n (ESTIMATE)$ Where,   a = intercept

b = gradient (slope)

The model parameters were approximated as:

	Visits	Nights	Expenditure
a.	2.521841	4.218551	5.837435b
b.	-0.47831	-0.44906	-0.42362

ATTACHMENT B

NVS DATA RELIABILITY

The results given in the NVS are based on a sample, rather than a census, of Australian residents. As with all sample surveys, the results are subject to sampling variability and therefore may differ from figures that would have been obtained if the entire Australian population had been included in the survey.

A measure of the possible degree of difference is given by the confidence interval, which indicates the extent to which an estimate might vary by chance from the true figure because only a sample of the population was included.

In Table A, the 95% confidence interval widths are given for a range of estimates available in the NVS. These confidence intervals are expressed as a percentage of the estimate. That is, there are approximately 19 chances in 20 that the true figure (which might be obtained from a census) is within the range identified by applying the figures in the table.

Table A   **Confidence intervals for NVS estimates at the 95% level**

Estimate ('000)	Overnight visitors	Visitor nights	Overnight expenditure	Day visitors	Day visitor expenditure	Overseas trips
20	>50	>100	>100	>100	>100	37.5
50	47.5	>100	>100	>50	>100	23.3
80	37.7	>100	>100	>50	>100	18.3
100	33.8	>100	>100	>50	>100	16.3
200	24.0	>50	>100	42.2	>100	11.3
300	19.6	>50	>100	34.6	>100	9.2
500	15.3	45.7	>100	26.9	>100	7.1
1 000	10.8	32.4	>100	19.1	>100	4.9
2 000	7.7	23.0	>100	13.5	>100	3.4
3 000	6.3	18.8	>100	11.1	>100	2.8
5 000	4.9	14.6	>100	8.6	>100	2.1
7 000	4.2	12.3	>100	7.3	>100	1.8
10 000	3.5	10.3	>100	6.1	>50	
20 000	2.5	7.3	>50	4.3	>50	
30 000	2.0	6.0	>50	3.6	>50	
50 000	1.6	4.6	>50	2.8	45.8	
70 000	1.3	3.9	49.5	2.3	39.8	
100 000	1.1	3.3	41.8	2.0	34.2	
140 000	0.9	2.8	35.7	1.7	29.7	
200 000	0.8	2.3	30.1	1.4	25.5	
500 000	0.5	1.5	19.5	0.9	17.3	
1 000 000		1.0	14.1		12.9	
2 000 000			10.1		9.6	
5 000 000			6.6		6.5	
10 000 000			4.7		4.9	
20 000 000			3.4		3.6	
50 000 000			2.2		2.5	

In Table A, the areas above the lines indicate estimates which have large confidence intervals (greater than 100% of the estimate above the solid line and between 50% and 100% of the estimate in the area between the dotted and solid lines). These estimates are subject to sampling variability which is too high for practical purposes and should be used with caution in analysis or in the reporting

of NVS estimates, as they have a large margin of error. The area below the dotted line contains estimates with smaller confidence intervals. That is, the estimates are closer to the values that would be obtained if the entire Australian population were interviewed.

ATTACHMENT C

INTERNATIONAL REGIONAL EXPENDITURE

The IVS provides information on travel activity and expenditure by international visitors. Information on expenditure by these visitors is only collected for whole trips; it is not regionally specific. In order to determine the impact that the visitor activity is having on a particular region, TRA uses a model-based approach to allocate visitor expenditure to the various tourism regions.

REGIONAL EXPENDITURE ALLOCATION METHODOLOGY

A very brief summary of the process by which expenditure by international visitors is allocated to regions is given below. A full explanation of expenditure allocation methods can be found in Carter and Collins (2005).

Two types of expenditure data are collected in the IVS:

- expenditure for the respondent's entire journey

- expenditure at a randomly selected location.

A 'location' is a more specific spatial unit than a Statistical Area Level 2 (SA2). For example, the location Bondi is in the 'Bondi - Tamarama - Bronte' SA2. A sub-sampling approach (selection of a single location for further study) is adopted as a starting point for regional expenditure estimation. This is because of the need for interviews to be done quickly and because it is unreasonable to expect an interviewee to remember expenditure at every stop. It is entirely feasible for a visitor to visit the same location more than once, but expenditure data are collected only if the randomly selected location has been visited only once.

Four major sub-components of total IVS expenditure are identified:

- expenditure on domestic airfares

- expenditure on travel packages

- expenditure on accommodation, food and beverages

- all other expenditure.

DOMESTIC AIRFARES

Wherever air travel is indicated, airfare expenditure is allocated equally between the departure region and the arrival region. In cases where there is no major airport in the departure and/or arrival region, the share of airfare expenditure is allocated to the nearest region with a major airport. If air travel is specified for the first leg of the trip, the departure airport is assumed to be the airport where the visitor arrived in Australia.

PACKAGE EXPENDITURE

A major part of any package for travel within Australia is taken up by airfares and other long distance travel fares. A series of studies by the ABS (1995, 1996), Australian Tourism Export Council (2000), Bureau of Tourism Research (Bonnet et al. 1994 and Skene, 1995) and Office of Economic and Statistical Research (2001) estimated the proportion of travel packages spent on things other than long distance fares to be between 26% and 35%. Based on these results, it has been decided that total package expenditure should be split with 30% being attributed to items other than long distance fares. The remaining 70% is assumed to be spent in the visitor's home country. The non-fare component is distributed among the regions using the iterative process.

ATTACHMENT C

EXPENDITURE ON ACCOMMODATION, FOOD AND BEVERAGES, AND OTHER EXPENDITURE

Expenditure on accommodation, food and beverages (AFB) and other expenditure is obtained by summing expenditure on the relevant items. The total expenditure for each item group is allocated to the regions by the iterative process.

During the iteration procedure, expenditure at the randomly selected location for which there is expenditure information is treated as a known value, and is held constant. The amount actually distributed among the remaining regions in the trip is known as net expenditure, which is equal to total expenditure minus random expenditure. If there is no expenditure at a randomly selected location (either by the interviewer failing to ask, or the interviewee failing to reply) net expenditure is equal to total expenditure.

THE ITERATIVE PROCEDURE

Steps in the iteration process are:

1. An initial regional cost indicator (average expenditure per night) is calculated for each region that has an expenditure sample. Three sources of data are used for this calculation: single region trips, random expenditure for one of the regions of two-region trips, and expenditure at a randomly selected location for multiple region (more than two regions) trips. National average expenditure is used for those regions where there is no expenditure sample.

2. For stops where there is no random expenditure, a preliminary estimate of expenditure at that stop in the trip is calculated by multiplying the cost indicator for the region at the stop by the length of stay at that stop. For stops where there is random expenditure, preliminary estimates of expenditure are left blank.

3. Non-blank preliminary estimates of expenditure are rescaled using the formula:

   $r = p * (E / R)$
   where
   r = rescaled value for this stop
   p = preliminary estimate for this stop
   E = reported total expenditure for the trip
   R = sum of preliminary estimates for the trip

4. For stops where there is random expenditure (that is, the preliminary estimate is blank), the rescaled value is set equal to the random expenditure. Rescaled values sum to the total reported expenditure for the trip.

5. New estimates of regional cost indicators are calculated by summing rescaled expenditure values for each region, and dividing this value by the total number of nights in each region.

6. If estimates of cost indicators for the current iteration differ from estimates of cost indicators from the previous iteration by less than an agreed amount the process is stopped, otherwise the procedure is repeated from step 2.

Rescaled values at the last iteration are the final estimates of expenditure at each stop.

At the end of the allocation process, expenditure on the four major expenditure components are estimated for every stop in each trip. Estimates of expenditure by state/territory and region are obtained by summing expenditure estimates at each stop.

It should be noted that expenditure by visitors who are in transit to another country is included in the expenditure estimates presented here.

ATTACHMENT C

NVS REGIONAL EXPENDITURE ALLOCATION METHODOLOGY

A brief summary of the process by which expenditure by domestic visitors is allocated to regions is given below. Further explanation of expenditure allocation methods can be found in *Travel expenditure by domestic and international visitors in Australia's regions, 1999–2010*, Tourism Research Australia, Canberra. Expenditure in the NVS is collected for the respondent's entire journey, not for individual stops. For both overnight and day visitors, information is collected on the following three types of expenditure:

- expenditure by respondent during the trip
- expenditure by respondent before or after the trip
- expenditure paid by employer or other who is not travelling.

For each of these three types of expenditure, information is collected on the amount spent on each of the following expenditure items:

- Packages
- Taxis (including to/from airport)
- Airline fares
- Organised tours/side trips
- Car hire costs (rental, leasing)
- Fuel (petrol, diesel)
- Vehicle maintenance or repairs
- Any other long distance transport costs (train, coach, ship etc)
- Any other local transport costs (bus, train, ferry, etc)
- Accommodation (not relevant for domestic day trips)
- Takeaways and restaurant meals
- Groceries etc for self-catering
- Alcohol, drinks (not already reported with food above)
- Shopping, gifts, souvenirs
- Entertainment, museums, movies, zoos etc
- Horse racing, gambling, casinos
- Conference fees
- Education, course fees
- Purchase of motor vehicles or any other major equipment
- Other (phone, postage, medical expenses, repairs, dry cleaning etc).

Expenditure on capital goods (for example, motor vehicles, property and office equipment) is not included in the published NVS estimates, as it is not regarded as tourism expenditure. All the other

expenditure items in the above list are included in the expenditure estimates for domestic day and overnight visitors.

OVERNIGHT TRIPS EXPENDITURE ALLOCATION PROCESS

As discussed above, expenditure information in the NVS is collected for entire trips, not for individual stops. A method has been developed by which this amount can be distributed over all the stops on the trip.

The process of allocating money spent on overnight trips begins with the identification of four major sub-components of total expenditure:

- expenditure attributable to the respondent's home region

- expenditure on airfares and other long distance fares

- expenditure attributable to the destination region or regions

- expenditure on long trips.

ATTACHMENT C

The 'home region' is the region where the survey respondent lives and home region expenditure, by definition, applies to a single region. In taking a trip a traveller may spend some money in the region where they live; this expenditure is identified as home region expenditure. For example, expenditure on taxi fares, fuel and groceries paid for before or after the trip are assigned wholly to the traveller's home region. Fuel expenses paid by someone other than the respondent and meals paid for before or after the trip are assigned to the home region in the proportion:

1 / (number of stops +1) – with the remainder going to destination regions.

Airfare and other long distance fare expenditure is allocated equally between the region where the journey started and where it ended. In the few cases where there is no major airport in the departure and/or arrival region (for example, the Central Coast of NSW or Melbourne East in Victoria), the share of airfare expenditure is allocated to the nearest region with a major airport. For trips where there is more than one destination, information on the transport mode used to get to each stopover is used to allocate airfares and other long distance transport fares to regions. For example, if a traveller uses air travel on three legs of their trip, one–third of the airfare expenditure is allocated to the departure and arrival regions for each leg.

Unpublished work by Tourism Research Australia has shown that, on average, about 60% of package expenditure is taken up by airfares and other long distance fares and the remaining 40% by other items like accommodation, food and car hire. The first step in the allocation of package expenditure is to split the expenditure into these proportions. The airfare and other long distance component are allocated together with other airfare expenditure and the remaining 40% is allocated by the iteration process.

In allocating package expenditure using the iteration method, the 'regional cost indicator' is average package expenditure per night. The end result of iteration is an estimate of package expenditure at each stop on each trip.

Destination region expenditure excludes expenditure on capital items and includes 40% of package expenditure. Sixteen expenditure items and package expenditure are grouped into nine major expenditure types, which are:

- Packages
- Local transport
- Entertainment
- Fuel
- Food
- Shopping
- Accommodation
- Conference fees
- Any other expenditure.

Each of these expenditure types is allocated by the iterative process.

For *expenditure on long trips* (trips with more than 21 stops) a single expenditure figure is collected for these trips, which is equal to total expenditure for the entire trip. This amount is allocated to regions in proportion to the nights spent at each stop.

ATTACHMENT C

THE ITERATIVE PROCESS

Overnight trips expenditure is allocated by iteration in three steps:

1. A regional cost indicator is calculated for each region that has an expenditure sample – national average expenditure used for those regions where there is no expenditure sample.
2. Preliminary expenditure estimates for each stop in the trip are calculated by multiplying the cost indicator for the region at the stop by the length of stay at that stop. Preliminary estimates of expenditure are rescaled using the formula:

   $r = p * (E / R)$

   where

r =	rescaled value for this stop
p =	preliminary estimate for this stop
E =	reported total expenditure for the trip
R =	sum of preliminary estimates for the trip

Rescaled values sum to the total reported expenditure for the trip.

3. New estimates of regional cost indicators are calculated by summing rescaled expenditure values for each region and dividing this number by the total nights spent in that region. If estimates of cost indicators for the current iteration differ from estimates of cost indicators from the previous iteration by less than an agreed amount the process is stopped, otherwise it is repeated from step 2.

Rescaled values at the last iteration are the final estimates of expenditure at each stop.

Experience has shown:

- Cost indicators for most regions reach a value close to their final value after four or five iterations
- Fifteen iterations are usually enough to get successive differences well below $1.00
- The process is not sensitive to initial estimates of cost indicators
- The main driver of the process is the nights spent at each stop.

DAY TRIPS EXPENDITURE ALLOCATION PROCESS

Expenditure allocation for day trips is similar in principle to expenditure allocation for overnight trips, but less complicated as each trip has just one destination region. All day trip expenditure is apportioned between home and destination regions using predetermined proportions, which are the same as those for overnight trips. Unlike the overnight trip analysis, there is no need to further apportion expenditure to multiple destination regions.

EXPLANATORY NOTES

MEASURING THE RELIABILITY OF WEIGHTED ESTIMATES

It is not possible to directly measure the reliability of each of the many weighted estimates which are derived from the IVS and NVS, so instead we must approximate it. The process used to derive the published reliability formulae and reliability tables for the IVS and NVS estimates measure the statistical variation of the data over a wide range of estimates. The recommended minimum useable estimates and sample sizes in the IVS and NVS are also derived from this process.

TERMINOLOGY

There are a number of words and phrases associated with the common measures of statistical reliability, some of which are confusing, and some of which are used interchangeably (and not always correctly).

- The **standard error** is the basic unit of sample error associated with a statistic (an estimate), and most other measures are derived from this unit.

- The **relative standard error** (RSE) is the standard error expressed as a percentage of the estimate.

- A **confidence interval** is an absolute measure of a range of estimates within which the "true" figure is likely to fall. Confidence intervals, as the name suggests, provide a measure of confidence expressed as a percentage, and the level of confidence required can be controlled by how many standard errors are used. Depending on the purpose of the research, different confidence levels are chosen. Medical trials often use a 99% confidence level (because people's lives may be at stake), which is 2.575 standard errors above and below the estimate. The most common confidence level for population-based surveys is 95%, and this is what is used for the NVS and IVS, and this uses 1.96 standard errors above and below the estimate.

- A **relative margin of error** (RMOE) is a confidence interval expressed as a percentage of the estimate. Like confidence intervals, the margin of error can be defined for any desired confidence level, but usually a level of 90%, 95% or 99% is chosen (typically 95%). Describing the reliability of an estimate in relative terms is beneficial in that it is easier for the mind to grasp. Because of this, it is common to see or hear the phrase "confidence interval of plus or minus … per cent". While this is not a strictly correct use of terminology, the meaning is usually clear.

FIRST ORDER AND SECOND ORDER WEIGHTED ESTIMATES

The NVS and IVS weighted estimates for trips or visits are **first-order** estimates, which means that they are simply the sum of the weights. The reliability of these estimates is determined by the sample size and the variation in the weights.

On the other hand, nights and expenditure are **second-order** estimates. The reliability of these estimates is the combination of the sample size, the variation within the weights **and** the variation within the nights or expenditure. For this reason, second order estimates will require more sample to achieve the same relative margin of error.

CHOOSING CUT-OFF VALUES

The common cut-off for a weighted estimate is a 95% relative margin of error of +/- 50%. When we extract figures from the IVS or NVS, there is no practical way to measure the RMOE, so we must use the reliability tables as a guide.

The IVS in 2013 has an average visitor weight of around 150. Statistical analysis suggests the cut-off for useable IVS visitor estimates is around 4000 visitors and a sample size of 30. However, if an estimate of this size had a significantly smaller sample, then we should suspect that the sample is likely to be unrepresentative of the population. For this reason, the sample size should always be taken into account when we are deciding which estimates to publish. TRA guidelines use an estimate of 5,000 visitors and a sample of 40. While this is slightly conservative, it does draw a line at a point where estimates are more likely to be reliable and takes some of the guess work out of choosing reliable estimates.

Note that for a data series which has a calculated cut-off of 30, this does not mean that an estimate which happens to have a sample of 30 is reliable, while another which has a sample of 29 is not reliable. Nothing magical happens at the cut-off, it is just that smaller samples are more likely to have problems. The cut-off simply tells us that samples or estimates above the cut-off are **more likely than not** to be within +/- 50% of the 'true' population figure, while those below the cut-off are more likely than not to be **outside** +/- 50% of the population figure.

However, arbitrary cut-offs are useful in practice, as it is not possible to examine every single number among a large set of tables, and decide if each one is OK. In addition, as the cut-off is at the bottom end of useable estimates, it is common practice to:

1. round up the estimated cut-off to the nearest 5 sample (if for no other reason than to make it easier to remember). TRA uses a publishing cut-off sample size of 40 per cell—this gives slightly more assurance around the estimate.

2. treat these estimates with caution and highlight affected cells/provide explanatory footnotes