Developing Survey Expansion Factors

Objective: To apply expansion factors to the results of a household travel survey and to apply trip rates to calculate total trips.

It is eighteen months later and the survey you designed in the first part of this case study has been completed. You are now given the assignment of taking the results and expanding them to reflect the total regional population. Let's assume your survey design was based on the following household type distribution and the actual number of completed household surveys shown below.

Target Sample Size for +/- 12.65% Error at 95% Confidence Level					
	Total	No vehicles	1 vehicle	2 vehicles	3+ vehicles
Total	2,000	489	507	508	496
1 person households	496	127	127	126	116
2 person households	505	124	127	127	126
3 person households	500	119	126	127	127
4+ person households	500	119	126	127	127

Actual Number of Households Surveyed by Household Type					
	Total	No vehicles	1 vehicle	2 vehicles	3+ vehicles
Total	1,910	451	503	520	436
1 person households	409	105	130	121	53
2 person households	511	121	131	134	125
3 person households	509	120	124	135	130
4+ person households	481	105	118	130	128

You will notice from these results that the household survey achieved its objectives for most of the household classifications. In a few cases, the number of completed surveys is significantly less than the target sample. For example, the number of completed surveys for one-person households with 3+ vehicles classification is much less than the sampling plan target. This is likely to reflect the relative difficulty of finding households of this type in the region. Other survey problems may reflect the difficulty of getting people of certain types to participate. They may have language barriers or they may not have a telephone. On the other end of the spectrum, they may feel they are too busy or don't want to be bothered with a survey.

Table 1:63: Number of Persons in Household (5) by Vehicles Available (6)						
	Total	No vehicles	1 vehicle	2 vehicles	3 vehicles	4+ vehicles
Total	380,070	21,995	123,660	166,710	50,215	17,490
1 person households	93,110	14,665	69,365	7,850	870	355
2 person households	125,155	3,750	34,350	75,285	10,070	1,695
3 person households	66,345	1,845	10,445	32,170	18,405	3,480
4+ person households	95,465	1,740	9,500	51,400	20,865	11,965

Calculating survey expansion factors

What are some reasons for calculating expansion factors?

One reason to calculate expansion factors is to use your survey results to calculate regional totals. Another reason is to look for sample bias in your survey. For example, if your sample design uses household size and vehicles available, you can examine the weighted survey results and compare regional totals for an independent variable (e.g., household income). You may find, that even by using the expansion factors, you may need other adjustments when calculating regional numbers to account for sample bias.

You will need to calculate an expansion factor for each of the 16 cells. For example, the expansion factor for 1-person households with no-vehicles is 140. This was calculated by dividing the total number of households from CTPP Table 1.63 (14,665) by the number of households from the survey (105), which is 140.

The survey has a category of 3+ vehicles, and the CTPP Table includes cells for 3 vehicles and 4+ vehicles. How would you calculate the expansion factor for 4+ person households with 4+ vehicles?

Since the survey data was collected based on a random sample of combined 3 and 4+ vehicle households, you may choose to split the survey responses by vehicle availability and calculate different expansion factors for each group. The results will not, however, have the same level of confidence or accuracy as the combined estimate. It is therefore better to combine the 3 and 4+ vehicle households into a single category and calculate a single expansion for the group. You may then choose to proportionally redistribute the total to the 3 and 4+ vehicle categories.

For example, the expansion factor for 1 person households with 3+ vehicles is calculated as:

$$(870 + 355) / 53 = 23.1$$

Here are the results for the 16-cell table.

Survey Expansion Factors by Household Type				
	No vehicles	1 vehicle	2 vehicles	3+ vehicles
1 person households	140	534	65	23
2 person households	31	262	562	94
3 person households	15	84	238	168
4+ person households	17	81	395	256

Calculating trips:
The survey estimated average daily home-based shopping trips per household, by household type, is shown below. Home-based work, home-based shopping, home-based other, and non-home-based trip rates are the types of attributes typically calculated from home interview surveys for travel demand forecasting models.

Home-based Shopping Trip Rates by Household Type				
	No vehicles	1 vehicle	2 vehicles	3+ vehicles
1 person households	0.80	1.20	1.24	1.22
2 person households	1.05	1.56	1.67	1.68
3 person households	1.20	1.69	2.03	2.50
4+ person households	1.25	1.90	2.45	2.74

To calculate total shopping trips by households with no vehicles (i.e., the first column), use the data from CTPP Table 1.63, and multiply the number of households by the number of trips. You do not need the survey expansion factors for this calculation.

	Number of Households	HBS Trip Rates	Home Based Shopping Trips
1 person households	14,665	0.80	11,732
2 person households	3,750	1.05	3, 938
3 person households	1,845	1.20	2,214
4+ person households	1,740	1.25	2,175
Total			20,059

By using rates, the trips for all household classifications can be calculated directly. This results in the table below. The households in the 4+ vehicles category are multiplied by the same trip rate at the households in the 3 vehicles category.

Total Home-based Shopping Trips per Day						
	Total	No vehicles	1 vehicle	2 vehicles	3 vehicles	4+ vehicles
Total	640,441	20,059	172,526	326,695	121,162	44,765
1 person households	105,765	11,732	83,238	9,734	1,061	433
2 person households	200,167	3,938	53,586	125,726	16,918	2,848
3 person households	131,184	2,214	17,652	65,305	46,013	8,700
4+ person households	203,325	2,175	18,050	125,930	57,170	32,784

Using the expansion factors:

If you did have data from survey counts, the expansion factors would be used to expand the counts to the regional population. If the following table represents the total number of workers in each household type, what would be the estimated number of regional workers who live in households with no available vehicles?

Number of Workers in the Survey by Household Type				
	No vehicles	1 vehicle	2 vehicles	3+ vehicles
1 person households	15	85	102	45
2 person households	76	102	199	235
3 person households	105	126	201	264
4+ person households	110	135	210	290

To estimate the number of workers in household with no vehicles, multiply the first column by the expansion factor:

	Number of No Vehicle Workers in the Sample	Expansion Factor	Number of No Vehicle Workers in the Region
1 person households	15	140	2100
2 person households	76	31	2356
3 person households	105	15	1575
4+ person households	110	17	1870
Total			7901

By applying the expansion factors to the whole table, the total number of workers in the region is estimated as:

Number of Workers in the Region by Household Type					
	Total	No vehicles	1 vehicle	2 vehicles	3+ vehicles
Total	493,021	7,901	93,582	249,529	142,009
1 person households	55,106	2,095	45,354	6,617	1,040
2 person households	163,235	2,355	26,746	112,016	22,118
3 person households	104,563	1,614	10,613	47,865	44,470
4+ person households	170,103	1,823	10,869	83,031	74,380

We can verify this estimate by comparing these numbers to other tables in CTPP 2000. For example, Table 1.65: Number of Workers in Household (6) by Vehicles Available (6) provides a distribution of workers by vehicle availability that we could compare to the totals in the table above.

Table 1.65: Number of Workers in Household (6) by Vehicles Available (6)						
	Total	No vehicles	1 vehicle	2 vehicles	3 vehicles	4+ vehicles
Total	380,070	21,995	123,660	166,710	50,215	17,490
No workers in household	81,080	15,960	44,240	18,690	1,810	380
1 worker in household	140,165	4,590	67,545	54,880	10,760	2,385
2 workers in household	128,430	1,265	10,890	87,160	23,780	5,340
3 workers in household	23,435	160	830	5,225	12,115	5,105
4+ workers in household	6,965	20	155	755	1,755	4,285

Notice that the numbers in this table are the number of households classified by number of workers and vehicles available. In order to convert the number of households to the total number of workers, you need to weight the households by the number of workers in the household. Assuming the 4+ workers category averages 4.2 workers per household, what is the total number of workers in households with no available vehicles?

Take the column of "no vehicles" and multiply each cell by the number of workers:

Household Category	Number of Households	Number of Workers	Number of Workers in Households with No Vehicles
1-worker	4,590	1	4,590
2-workers	1,265	2	2,530
3-workers	160	3	480
4+ workers	20	4.2	84
Total			7,684

If you complete the calculations and compare the results to the expanded total from the household survey, you get the following result:

Number of Workers in the Region by Household Type					
	Total	No vehicles	1 vehicle	2 vehicles	3+ vehicles
Survey Results	493021	7,901	93,582	249,529	142,009
CTPP 2000	496609	7,684	92,466	248,046	148,413
% Difference	-0.7%	2.8%	1.2%	0.6%	-4.3%

These results show that the survey slightly over estimated the number of workers in no, 1, and 2 vehicle households and under estimated workers in 3+ vehicle households. The overall total is slightly less, but very close. In general, most planners would be very pleased with these results. Congratulations, your survey was a big success!

So what would you do if the results did not match very closely? One approach might be to re-weight the survey records by number of workers in the household. In this approach you would compare the distribution of workers by vehicles available found in CTPP 2000 Table 1.65 to the distribution of workers by vehicles available found in the survey. If the distribution is considerably different, you can apply a correction weight to the household records where the differences exist to increase or decrease the contribution of those household records to the calculated statistics. For example, if you find that the expanded survey has 20 percent more households with 2 workers and 2 vehicles than the CTPP 2000 table identified, you can apply a weighting factor of 0.8 to each 2 worker-2 vehicle household record in the survey and recalculate the trip rates using the record weights.

Note there is one more calculation that you should always perform on the survey results to check how well the overall accuracy targets were met and what the actual error in the estimate is. Calculate the mean and standard deviation for each cell based on the survey results and then use the Coefficient of Variation formula and the actual sample size to calculate the percent error. The formula is shown below:

$$\text{Error} = (\text{Standard Deviation} * CF) / (\text{Mean} * (\text{Sample Size})^{1/2})$$

where:

Error is the maximum error in the estimate

CF is the confidence factor based on the target confidence level
CF = 1.645 for a 90 percent confidence level
CF = 1.960 for a 95 percent confidence level
CF = 2.576 for a 99 percent confidence level

If you apply this formula to the home-based shopping results for 4+ person households with 2 vehicles available (given a standard deviation of 1.93), you will get the following results:

$(1.93 * 1.96) / (2.45 * (130)^{1/2}) = 0.1354 = 13.54\%$

Note that this is worse than the 12.65 percent error assumed in the sample size estimation. Since our sample was based on total person trips and this calculation is based on home-based shopping trips, it is logical to expect the error of a subordinate statistic (i.e., one that is a subset of the target statistic) to be less accurate than the target statistic. The impact of this calculation is that we could say with 95 percent confidence that the average trip rate for home-based shopping trips for 4+ person households with 2 vehicles available is between 2.12 and 2.78 (+/- 13.54%) person trips per day. By applying this calculation to each statistic and each cell, the user can determine the percent error of each statistic. It would not be correct to assume the percent error used in the sample size calculation applies to all statistics calculated based on the survey results. Even the percent error for the target statistic (i.e., total person trips) should be recalculated based on the actual results of the survey. This calculation will verify how well your survey design captured the actual variability of the population.