

Task 1.4 Household Survey Processing (Auckland example)

Data Acceptance

The raw HTS data was received from the survey agency as 5 separate files linked by household, person and vehicle IDs. The number of records in the final version of the data received from the survey agency, TUTI, is given in Table 1.

Table 1: HTS Data Received

File	Weekday	Saturday
Household File	5,221	808
Person File	15,481	2,364
Vehicle File	10,153	1,583
Stop File	51,407	6,982
Trip File	46,677	6,759

The trip file provided was not part of the survey brief and was not used; as part of the data processing trip linking of stops was undertaken to produce a new trip file using different criteria from those applied by the survey agency.

■ *Survey Sample*

The survey was to cover a sample of 6,000 responding households, but instead the weekday and weekend household interviews totalled 6,000, so with the weekend households also surveyed during the week, the number of households interviewed was only 5,221. The loss of weekday sample was considered acceptable in terms of the model development requirements.

■ *Responding Households*

A responding household in the survey was defined as:

- at least 50% of travel diaries provided;
- no missing data on key items of information (see MIC requirements below);
- no more than 10% of missing responses for all variables other than those defined as key items, excluding missing diaries.

All households in the database have at least 50% of travel diaries provided; 6% of households had some missing travel diaries (in most cases one diary was missing), hence no records have been rejected as a result of this requirement.

One household has more than 10% missing responses (10.4%). This is a household in which no travel took place on the travel day and the data on employment type and occupation is missing from the person file. Hence the missing data will have no effect on the trip file, so has been accepted.

Across the entire sample of households, the requirements were that there be:

- no more than 5% of diaries missing from members of households;
- no more than 10% of diaries missing from members of households in any one of the sampling sectors;
- no missing data on key items of information (see MIC requirements below);
- no more than 3% of missing responses for all variables other than those defined as key items (excluding the personal income question).

As shown in Table 2 the survey as a whole has less than 5% missing travel diaries and all sampling sectors have less than 10%.

Table 2 Missing Travel Diaries by Sampling Sector

Sampling Sector	% Missing Diaries
1	1.0%
2	1.7%
3	2.1%
4	2.4%
5	2.2%
6	2.1%
7	2.3%
8	3.3%
9	4.3%
10	2.7%
11	1.0%
Whole Survey	2.4%

Missing responses made up less than 3% for all variables, except for the four employment-type and occupation questions in the Person file. For these questions, 7.2% to 8.0% of coded responses are missing, but given the nature of this question the data has been accepted.

■ *Key Items*

The minimum information content (MIC) requirements were that interviews would be rejected if they do not contain the following key items:

- Household Data
 - Household Address;
 - Household size;
 - Number of passenger vehicles.
- Person Data
 - Age;
 - Employment status

- Stops and Trips:
 - Location of day start (for first trip only);
 - Time start trip stage;
 - Type of place at end of trip stage;
 - Address at end of trip stage;
 - Trip stage purpose;
 - Trip stage mode of travel;
 - Time of arrival at next stop.

The only data item which did not fully comply was “Location at the day start” for which 205 records (0.35%) were missing, but for which x-y co-ordinates and meshblocks were allocated, presumably based on the location of the household. This exception is considered acceptable and the data complies with the MIC.

Hence all records received have been accepted and the final databases contain the numbers of records as given in Table 1.

Range and Logic Checks

Comprehensive checking of the raw HTS data was undertaken to ensure sensibility of the data received which, in conjunction with the following data processing stages, lead to revised versions of the data being provided. The ARC checked the geo-coding and a sample of interview forms compared with the databases. The former lead to households and trip ends being coded to 2006 meshblocks as well as x-y co-ordinates. The latter led to some revisions in the data.

Checks have been carried out with regard to:

- Consistency within each database;
- Consistency between the databases;
- Consistency with the data processing specification.

Trip Linking

This process links trip stages into trips on the basis of a change in mode between stages.

The separate records of stops in the stop file become single trip records in the trip file and the origin and destination purposes of the trips are those of the first and last stages respectively. As noted below, some records were identified where the trip starts or ends with a change of mode; in these cases the trip purpose has been determined from the destination place and where this was not possible they were coded as “other”.

The mode of linked trips has been defined as noted below and other trip-stage data (such as mode, distance, time) have been attached to the trip as required.

Changes of mode defined where trip stages were linked into single trips, and three categories were considered and implemented in this order:

- PT trip linking;
- Car driving and parking trip linking; and
- Escort trip linking.

- *PT Trip Linking*

The main criterion for a linking trip stages involving PT is that the origin and/or destination purpose should be “getting on or off PT” and the main mode of travel being PT was considered as a check.

Variations in the duration between getting on or off PT and the previous or next stage of a trip were investigated which showed times of more than an hour and up to eight hours. These were identified as incidences of “getting on or off PT” being recorded as the origin purpose of the first stage of a trip or the destination purpose of or last stage of a trip. These seemed to occur where the walk distance before or after the PT stage were short and possibly off-street, such as within a shopping complex or a school. Different duration thresholds were tested for each PT mode before deciding on the most appropriate for each, taking into account generally the relative frequencies of each.

The mode of the linked trips was the same as the PT mode and for multi-leg PT trips, the modal hierarchy applied.

The final criteria applied were:

- Origin or destination purpose is getting on or off PT, and
- The main mode of travel for a stage is PT, and
- The duration between trip stages is less than 15 minutes for bus and less than 30 minutes for ferry and train.

- *Car Parking Trip Linking*

Trip stages involving parking or un-parking a vehicle and walking were linked into single trips. The mode of travel (driver or passenger) was considered as a check and for linking to occur there needed to be a walk stage before un-parking or after parking. This was necessary due to parking or un-parking being recorded as the purpose for the first or final stage of a trip (much the same as with PT as discussed above).

The mode of the linked trips was the same as the car leg, that is, driver or passenger.

Hence, the criteria applied to identify this category of trip linking were

- the origin or destination purpose is “parking or un-parking a vehicle”, and
- the mode of the car-based stage is vehicle driver or passenger, and
- the parking or un-parking of a vehicle is followed by a walk stage or vice versa.

- *Escort Trip Linking*

As identified during the specification of escort trip purposes there is one case where escort-related trip stages are linked. This is where a driver travels to another household in order to pick up someone from that household and then continues on to the final destination and vice versa.

The mode of the linked trips was car driver.

In this case the stages were linked if the following conditions were satisfied:

- the destination purpose is “accompanying someone or dropping someone off or picking someone up”,
 - the destination place is “someone else’s home”,
 - there is a change in non-household occupancy, and
 - the person is the driver of the vehicle.
- *Resulting Trips*

The 51,407 weekday and 6,982 weekend records in the Stop file were reduced to 46,677 weekday and 6,759 weekend trips in the Trip file.

Expansion to Census Data

The survey data has been expanded to the standard meshblock-based 2006 Census data released by Statistics NZ. A separate order of data specifically designed for the expansion and bias correction has been placed with Statistics NZ, but at this time it is unknown when this will be available.

■ *Household Expansion*

The expansion of households involved the following steps:

- The ART3 transport zones were aggregated into expansion areas of reasonably uniform household samples and with a minimum sample of around 30 household interviews; this resulted in 148 expansion areas are shown in Figure 1 and the distribution of the samples in each are given in Figure 2;
- Households by separate dwelling households and other households (apartments, flats, terrace houses, etc) were extracted from the Census data and allocated to each expansion area;
- The interviews were expanded by the two household types and the ranges of expansion factors and samples reviewed, looking particularly for high factors and low or zero samples; this highlighted the low sampling – zero in some expansion areas - and hence high expansion factors of “other households” (attached dwellings) compared with separate dwellings and an alternative approach was developed to account for this;
- Separate dwellings in the HTS were expanded to the separate dwellings in the Census directly;
- A bias correction factor of 4.621 was calculated and applied to other (attached) households along with the expansion factors for separate dwellings.

The separate dwellings expansion is:

$$\begin{aligned}
 C_S &= E_S \\
 &= EF_S S_S \\
 EF_S &= \frac{C_S}{S_S}
 \end{aligned}$$

The bias factor is:

$$\begin{aligned}
 \sum C_A &= \sum E_A \\
 &= \sum [S_A EF_A] \\
 &= \sum \left[S_A \frac{C_S}{S_S} BF_A \right] \\
 &= BF_A \sum \left[S_A \frac{C_S}{S_S} \right] \\
 BF_A &= \frac{\sum C_A}{\sum \left[S_A \frac{C_S}{S_S} \right]}
 \end{aligned}$$

The other (attached) households expansion is:

$$\begin{aligned}
 C_A &= E_A \\
 &= EF_A S_A \\
 EF_A &= f(EF_S) \\
 &= \frac{C_S}{S_S} BF_A
 \end{aligned}$$

Where: C = Census

E = Expanded

S = Survey

EF = Expansion Factor

BF = Bias factor

Subscripts A and S indicate Attached and Separate Dwellings.

The average separate dwelling expansion factor is 79.8 and their distribution is shown in Figure 3. For other (attached) households the average expansion factor, including bias correction, is 310.8, and the distribution is shown in Figure 4. For the latter, there are some examples of very high factors which are the result of the low sampling of other households.

Figure 1 HTS Expansion Areas

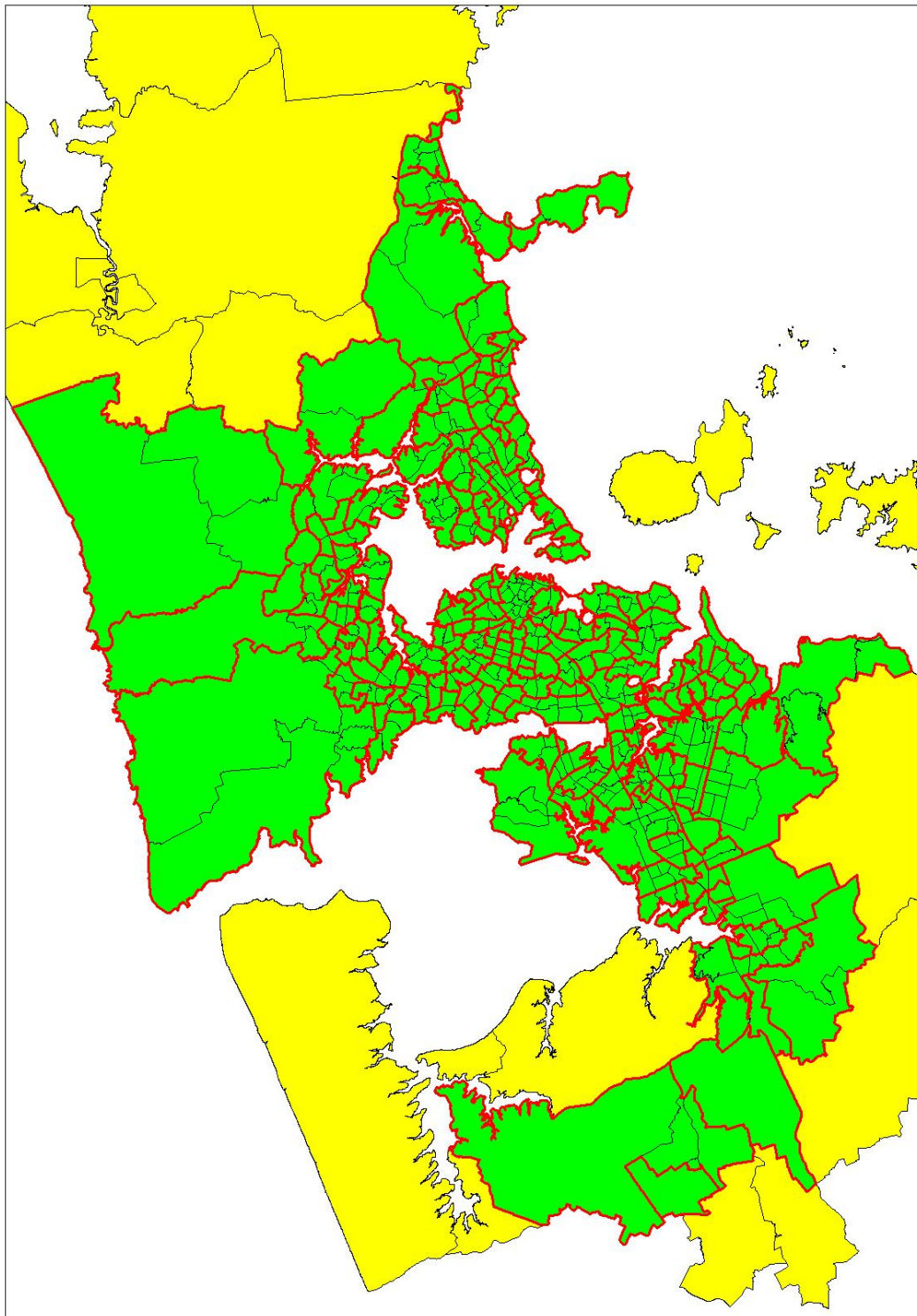


Figure 2 Distribution of Sample Size in Expansion Areas

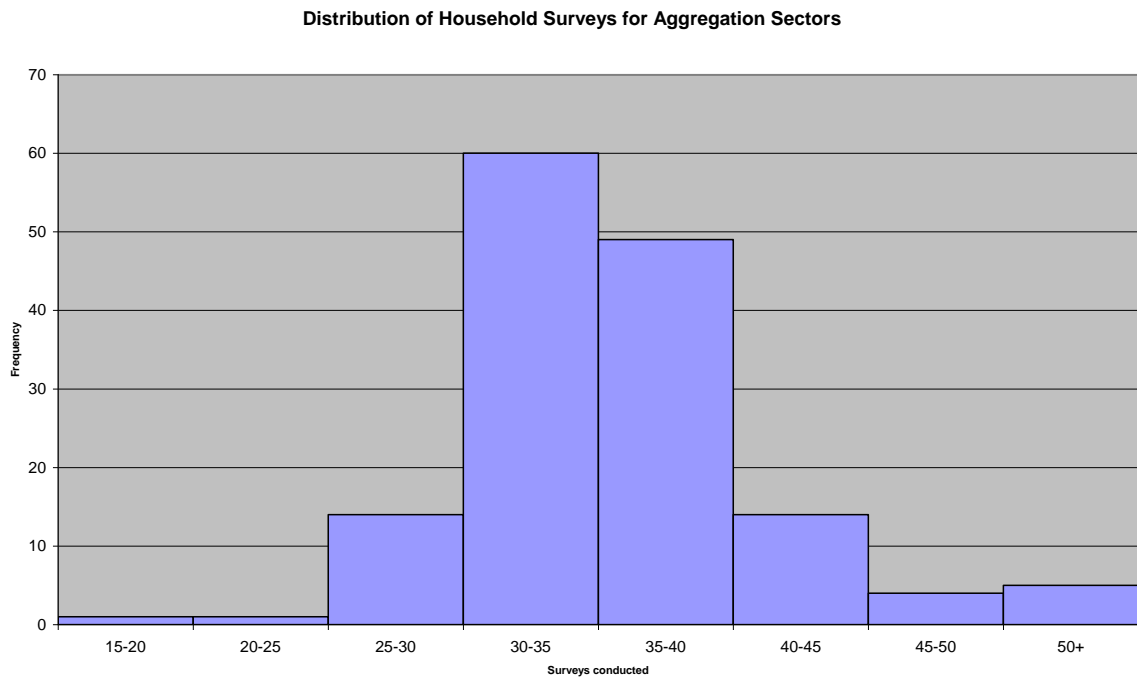


Figure 3 Distribution of Expansion Factors for Separate Dwellings

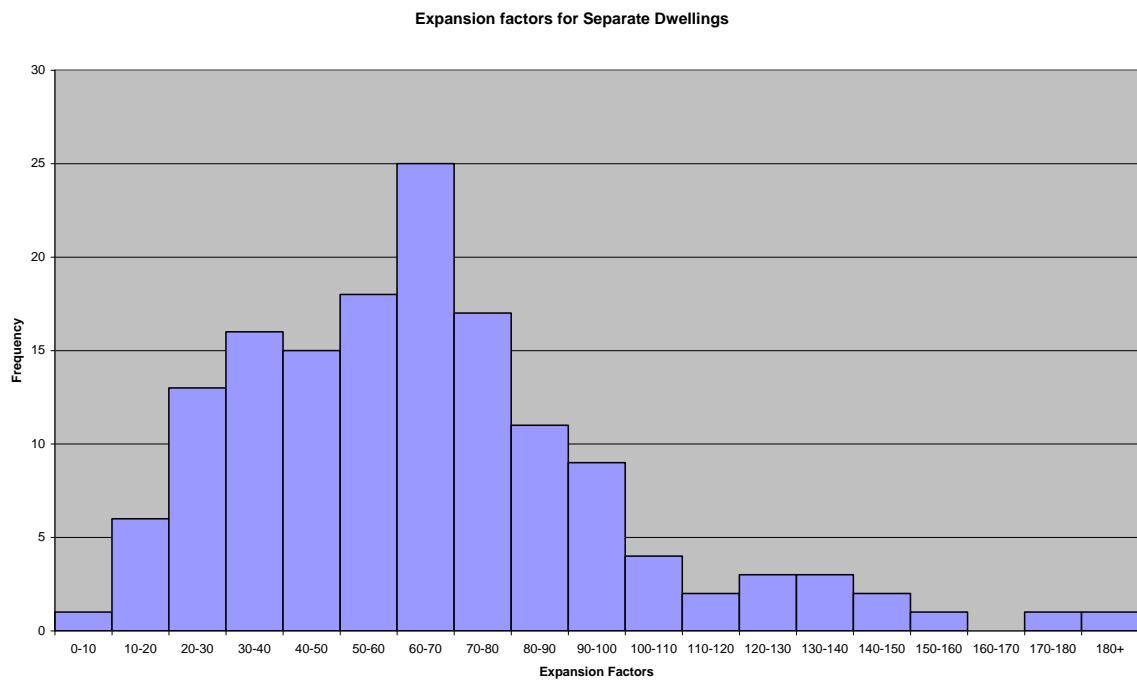
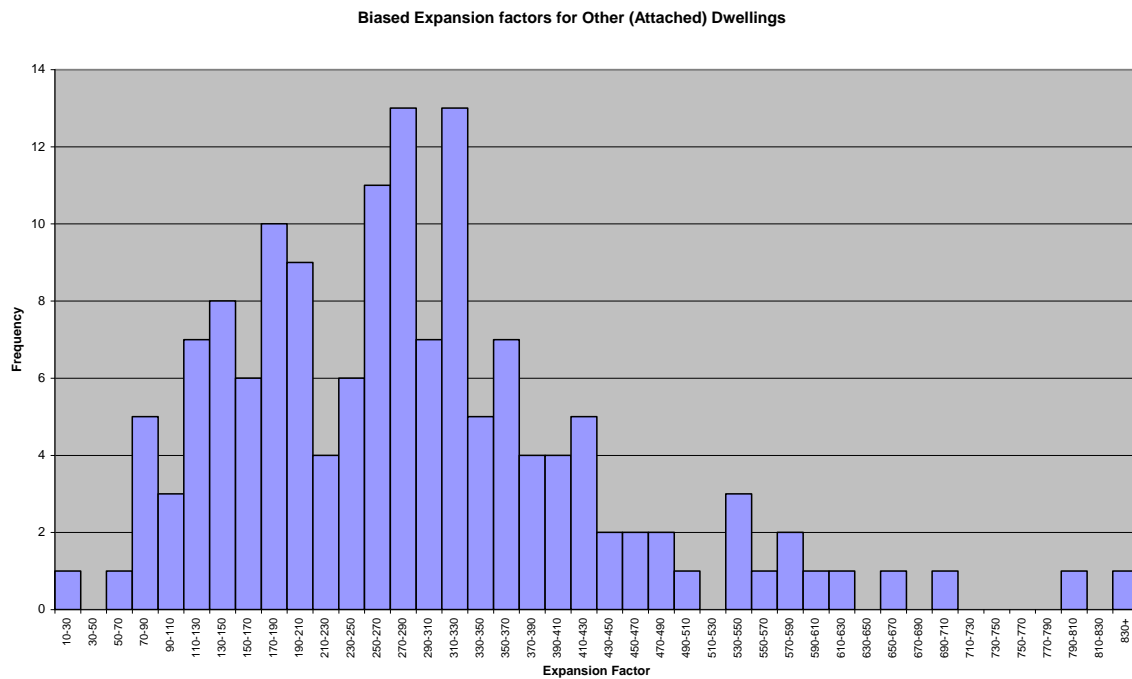


Figure 4 Distribution of Expansion Factors for Other (Attached) Dwellings



- **Person and Trip Expansion**

The expansion of persons and trips used the household expansion factors plus an additional correction factor to account for missing travel diaries. For this the following are calculated:

- the number of persons in each expansion area 'c' in age/working status person categories 'a', N_{ca} ,
- the number of missing diaries in each area by person category, n_{ca} ,
- the correction factor to be applied to all trip records: $1/(1 - n_{ca}/N_{ca})$.

The person categories used are aggregations of the person types (PERTYPE):

- Dependent children, aged 5-17
- Working adults
- Other adults, not retired
- Retired adults

The distributions of these correction factors for each category are shown in Figure 5, Figure 6, Figure 7, and Figure 8.

Figure 5 Distribution of Correction Factor (aged 5-17)

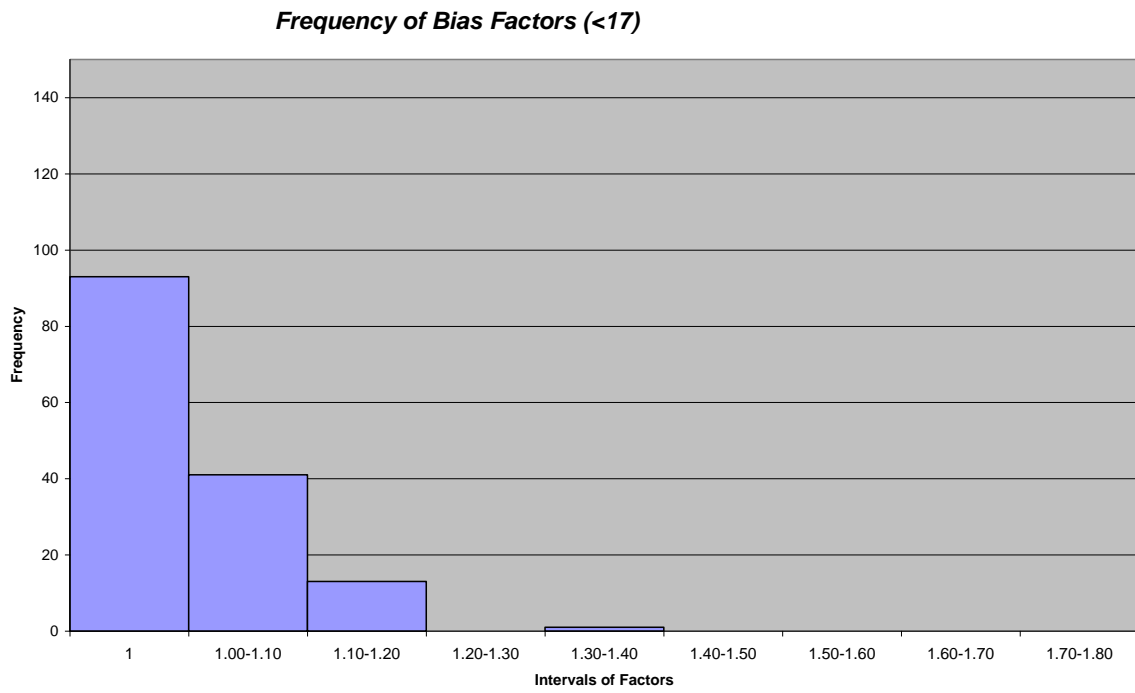


Figure 6 Distribution of Correction Factor (full time adults)

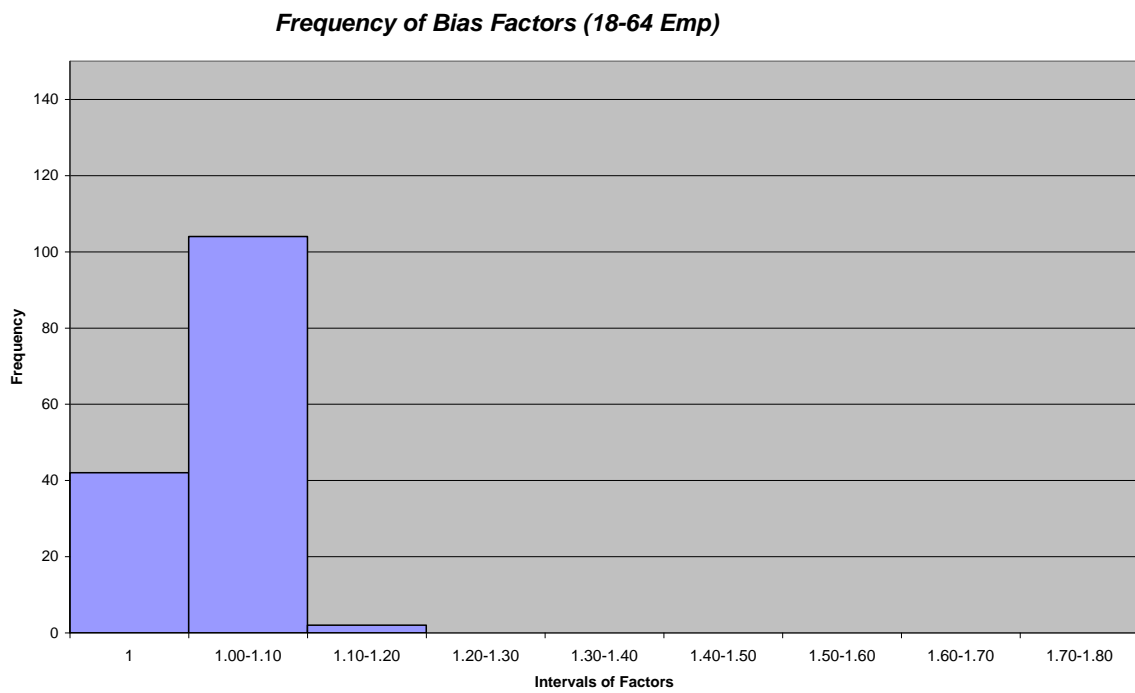


Figure 7 Distribution of Correction Factor (other adults, not retired)

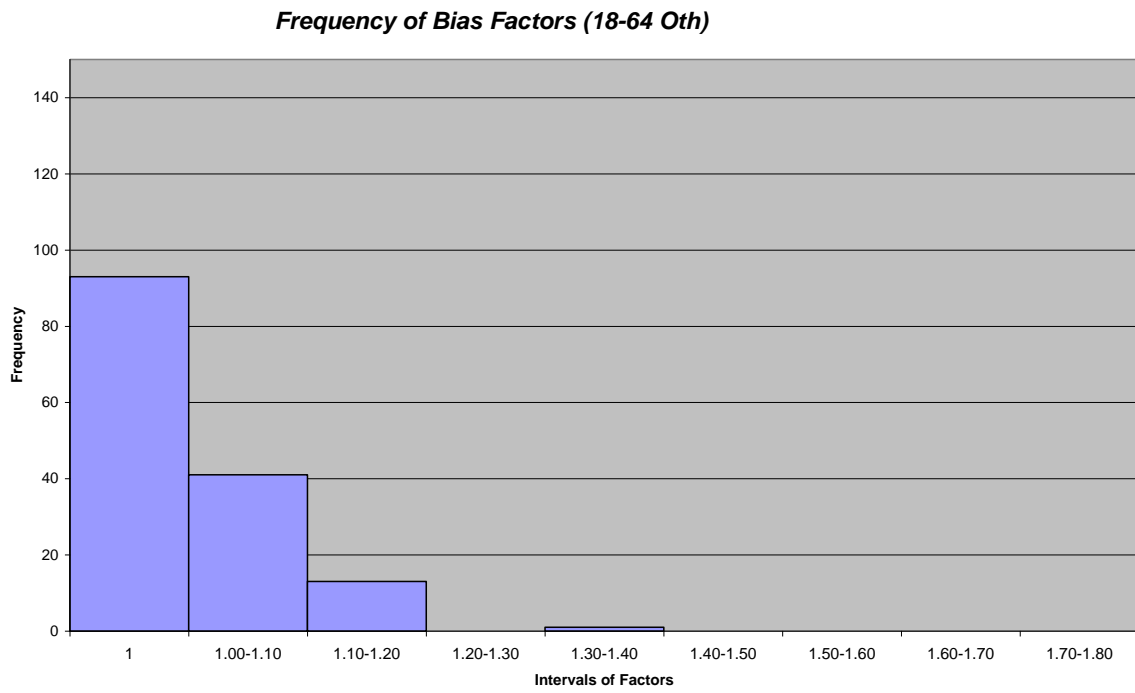
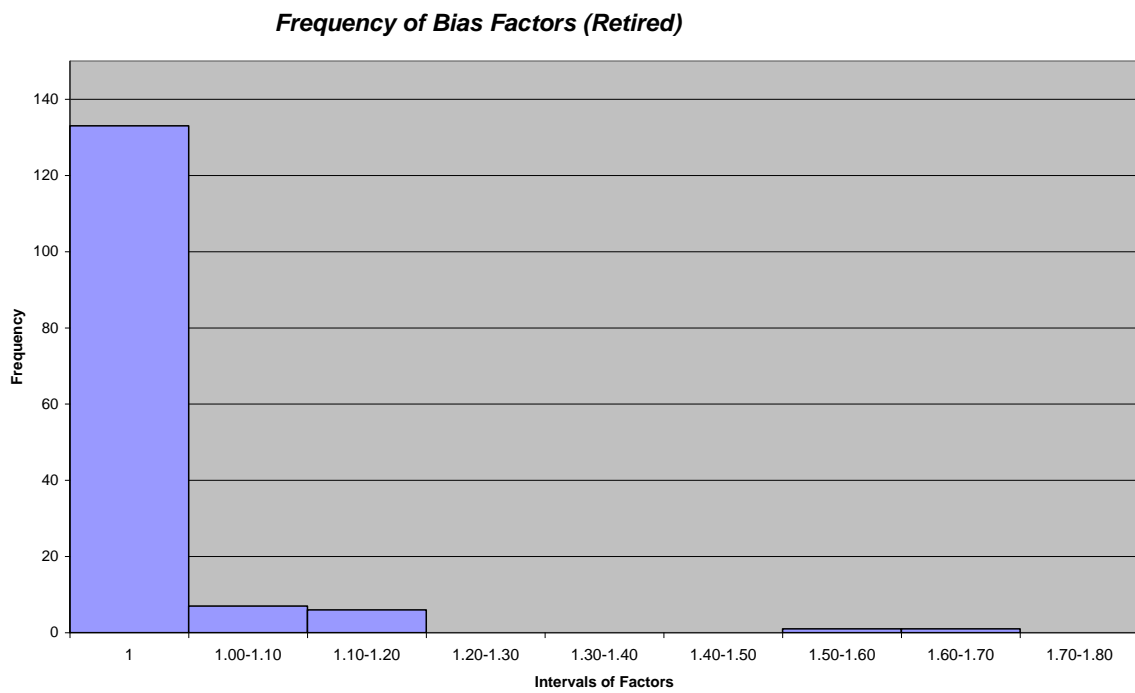


Figure 8 Distribution of Correction Factor (retired)



Final Database

■ *Documentation*

The final processed HTS data was converted from spreadsheets (as provided by TUTI) into Microsoft Access databases:

- Household file
- Person file
- Vehicle file
- Stop file
- Trip file

Standard coding common to all surveys was applied to the databases in line with the coding compatibility specification. The codes and their descriptions applicable to the HTS are given in Table 3.

Table 3– HTS Common Synthesised Codes

Code	Description
PERIOD	Time period
TRIPPURP	Trip purpose
PROD	Trip Direction (Production/Attraction)
MODE1	Mode of first leg of trip
MODE2	Mode of second leg of trip
MODE3	Mode of third leg of trip
MMODE	Main mode of transport for surveyed trip
CARAV	Level of car availability
PERTYPE	Type of person
WORKFP	Full time/part time worker
WORKSE	Socio-economic group
EMPTYPE	Type of employment
WORKARR	Work arrangements
HHTYPE	Household type
OZONE	ART3 Zone of origin
DZONE	ART3 Zone of destination

As well as the above codes most of the original information from the raw databases has been retained in the final databases.

■ *Verification*

Summary statistics from the expanded HTS databases are given in Table 4, Table 5, Table 6, and Table 7. Table 4 is a summary of the expanded households by type and TA (for the parts covered by the survey) compared with Census data.

Table 4 Expanded HTS vs Census Households

Area	Separate Dwellings		Other Households		Total Households	
	HTS	Census	HTS	Census	HTS	Census
Rodney	17,826	17,826	3,165	2,548	20,374	20,991
NorthShore	55,674	55,674	17,154	13,156	68,830	72,828
Waitakere	52,765	52,765	9,608	8,134	60,899	62,373
Auckland	85,733	85,733	56,026	58,807	144,540	141,759
Manukau	76,947	76,947	16,707	21,153	98,100	93,654
Papakura	12,737	12,737	2,170	1,265	14,002	14,907
Franklin	7,189	7,189	935	702	7,891	8,124
Total	308,871	308,871	105,765	105,765	414,636	414,636

Table 5 Expanded HTS Household Data

Household Size	HTS
1	75,567
2	133,374
3	81,770
4	76,592
5	31,507
6	12,956
7	1,989
8	545
9	336
Household Licence Count	HTS
0	3,149
1	83,711
2	122,942
3	83,729
4	72,853
5	30,915
6	14,616
7	2,720

Table 6 Expanded HTS Person Data

Sex	Persons
Male	577,215
Female	623,471
Employment	Persons
Full Time	496,947
Part Time	113,831
Casual	36,078
Studying	Persons
Primary School	116,958
Secondary School	110,327
Full Time Tertiary	54,550
Part Time Tertiary	26,399
Other	14,751

Table 7 Expanded HTS Trip Data

Period	HTS Trips		
AM Peak	820,864		
Interpeak	1,059,762		
SchoolPeak	456,146		
PM Peak	611,619		
Off-Peak	560,404		
Trip Purpose	Non-Escort	Escort	Total HTS Trips
HB Work	663,767	23,804	687,571
HB Education	405,072	81,414	486,486
HB Shop	321,732	934	322,666
HB Social	140,760	5,847	146,607
HB Other	490,521	228,617	719,138
EB	506,501	8,231	514,732
NHB Other	444,647	186,918	631,565
Production / Attraction	HTS Trips		
From Home	1,251,386		
To Home	1,205,356		
Non-Home-Based	1,058,138		
Main Mode	HTS Trips		
Car Driver	1,958,177		
Car passenger	655,747		
LCV driver	170,456		
LCV passenger	58,818		
Motorcycle driver/ passenger	8,145		
Taxi passenger	17,137		
Bus	107,438		
Train	22,281		
Ferry	6,559		
School bus passenger	36,334		
Walk	415,477		
Cycle	24,339		
Trucks (MCV and HCV)	20,183		