Task 1.6 Public Transport Survey Editing and Processing

The processes described below may be undertaken within a relational data base, many tasks being done simultaneously.

Inputs

Rail, ferry and bus survey questionnaire database.

Interviewer records

ETM data

Processing

The bus expansion is specified first, followed by rail and ferry, and then separate databases are combined.

1) For bus, service-specific expansion factors are to be appended to the trip records as follows:
   - pre-process the questionnaire database, using the questionnaire number in conjunction with interviewer records if necessary, to allocate a point of survey/boarding to each record, a time of survey (according to the time periods recorded by the interviewer) and a point of alighting; this should establish a unique correspondence between the count and questionnaire data;
   - process the counts; these distinguish in-scope from out-of scope (children) and boarding from alighting;
   - pre-expand the questionnaires to in-scope boarding counts by individual service and boarding stop; check the size of expansion factors and consider aggregating over stops for large factors[1];
   - append the expansion factors to interview records.
2) To investigate biases related to short trips and schoolchildren:
   - check the validity of the expansion by tabulating expanded questionnaires by alighting stop (presume feasible) and comparing with in-scope (ie adult) alighting count; analyse by (increasing) stop number towards main alighting points; look for evidence that alighting at intermediate points is under-estimated;
   - on the basis of this analysis infer a single set of short trip bias correction factors to be applied across the full data set, then re-expand using the factors[2];
   - extend the expansion to allow for accompanied children (5 years or older) recorded on the questionnaires;
   - append the expansion factors to interview records;

---

[1] In this process new issues may arise such as stations/routes with no achieved questionnaires in a time period or with more questionnaires than the counts, which will require adjustment to the expansion process.

[2] The factor would estimate the relative likelihood of questionnaire return for short trips and, in the re-expansion, this bias factor would be applied prior to expanding to the boarding counts.

- compute the difference between the expanded questionnaire boarding and alighting flows and counts of all passengers (allowing for under 5s); this difference is, presumably, unaccompanied children; how these data are processed further is handled in a separate task (refer forward).

3) The ETM data for expanding to the survey period is expected to comprise weekday bus passengers by route by direction by hour for the 5 week survey period; it may be available by ticket type, in which case the following process may require modification:
   - aggregate counts and interviews by hour by direction to the bus route corridors (combining services which form a corridor) as this is the sampling frame for the services;
   - tabulate the interview samples and counts for each corridor by hour; determine whether/where to aggregate hours for the expansion (allow for need to retain peak/off-peak differentials); aggregate time periods where the number of interviews is too few in the basic time periods to reliably expand (eg expansion factors would be too big);
   - compute expansion factors as ratio of counts to expanded questionnaire interview sample for each corridor, direction and time period;
   - append the expansion factors to interview records;
   - verify the process by tabulating expanded data and comparing with counts.
4) Expansion to 24 hours is done by adding to the trip records 24-hour factors by direction derived from ETM data for each corridor. It is proposed to consider making this factor purpose-specific, based on analysis of the household survey.
5) Expansion to a representative period can be considered (based on long-period ETM data), but is not currently expected to be needed.
6) Expansion of the train and ferry interviews will follow similar, though somewhat simpler, procedures. But, additionally, as they are in a single direction of travel trip reversal is required. This will be done before 24 hour expansion, as follows:
   - the survey records are duplicated, with the O and D zones and purposes are swapped;
   - for the interview direction, the time of travel is the time of interview;
   - for the return trip, the time of travel is that provided on the questionnaire by the respondent[3];
   - for some questionnaires the return journey time will not be provided, either by omission or because the trip is 1-way; an appropriate process is to be developed, for example:
     - the reversed interviews for which a time of travel is available can be expanded by hour to the reverse counts;
     - if the irreversible trips have different characteristics to the others then extra bias factors may be applied to the reversed trips, based on the interview-direction

[3] Check whether this needs to be adjusted in any way (to which point in the journey it refers).

time period and trip purpose, to account for the specific characteristics of the trips which cannot be reversed.

7) The surveys on buses and trains are in principle subject to double-counting[4] if more than one train or more than bus was used on the observed trips (as can be approximately determined from the access/egress mode data on the questionnaire):
   - tabulate the number of double/triple-counted trips of all possible types in each data set and compare the respective volumes: if double/triple-counted the volumes of the double-counted trips in each affected data set should be broadly similar; it may also be useful to tabulate and compare their geographic distributions;
   - an additional expansion factor of 0.5 (or 0.33) should be appended to any questionnaire in which the respondent used the same mode for access or egress (or both); this factor will need to be reconsidered if the affected trips in each data set are dissimilar.
8) The final task is to combine the bus, rail and ferry databases into one.

Outputs

Expanded trip database.

Note.

[4] This refers to within-mode double-counting. In a later task we look at double-counting between modes.