Session 6.2: Trainee Notes

Introduction to Quantitative Research Methods

Contents

Learning Objective

After completing this session, you will:

- Understand how to define the research question and hypotheses
- Be able to identify a population sample and remove bias from the sample
- Know how to undertake data collection by questionnaire survey
- Be able to design the following types of survey:
    - Key informant questionnaire
    - Household questionnaire
    - Transport operator questionnaire
    - Travel diary
- Weight the sample to redistribute the population of respondents more appropriately
- Understand ethical considerations of undertaking research with human participants

1. Things to consider

The following headings indicate issues that should be addressed when considering data collection and experimental design.

Stating the Objectives:

State clearly the objectives of why the data is to be collected and what research questions the results are expected to answer.

Define the Population about which Inferences are to be made:

The population of interest shall be precisely defined; this may be items to be measured or tested, vehicles, drivers, pedestrians or whatever.

Selecting the Sample:

Select the sample to be measured, tested, or interviewed from the defined population. Define a sampling methodology, i.e. simple random, stratified, cluster, systematic or quota sample.

The sample to be selected should be representative of the defined population, such that no known biases are introduced.

Calculating the Sample Size

How many people do you need to survey? Do the survey results need to be statistically significant? What is the sample size as a proportion of the population? (1%, 10% more?). Think about what results you are trying to achieve in relation to your research question and then consider what sample size is appropriate to respond to your hypotheses. Also consider your budget for conducting the survey and how this affects sample size.

Define the data Collection procedure

What research methods do you need to use to achieve the results to answer your original research question? Qualitative/quantitative research methods? Combination of both for added rigour? How many survey questionnaires do you require? If your research question relates specifically to transport service operators, you will need to survey operators, but you may want to collect the views of transport service users as well in order to capture a more rounded response. How much time do you have to collect the data? Prioritise your data collection requirements.

Is there bias in the process?

Identify the risks of undertaking a survey and how to mitigate against those risks. Is your sample broadly representative of the population? Go through the bias checklist (see slide on Bias later in this presentation) to ensure you reduce the risk of bias in your sample. If in doubt aim for a random sample and pilot the survey following checklist interviews.

2. Sampling

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.

Definitions

Population: The set of units that you hope to generalise to

e.g. people in Berkshire

Accessible population: The population that is accessible to you

e.g. people in the phone book

Sampling frame: Procedure for sampling accessible population (this may be a list)

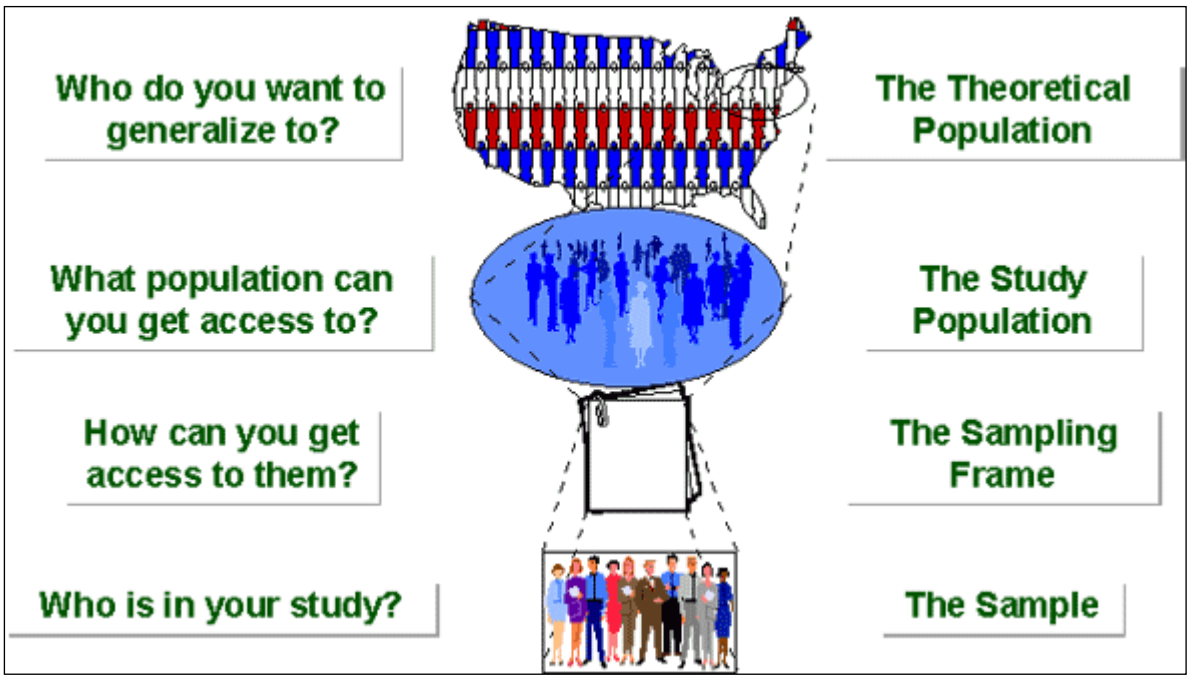
e.g. the phone book

Sample: The group of units who you select to be in your sample

e.g. a simple random sample from the phone book

NB the group that you select and the final group you get may not be the same

- Non-response
- Drop outs
- Cannot contact



http://www.socialresearchmethods.net/kb/sampterm.php

In identifying a sample, there is the possibility of introducing systematic error or bias at every stage.

- You may not have access to the whole population
- You may not have a complete and accurate enumeration or sampling frame from which to select
- You may not draw the sample correctly or accurately
- They may not all come and they may not all stay.

Simple Random sampling

- Each unit has an equal chance of being selected.
- Suitable for a small sample with a complete sampling frame.

Procedure

1. Select your population and sampling frame
2. Give each unit in your population a unique number
3. Calculate/Guess your sample size
4. Retrieve that many random numbers from a random number table or generate random numbers in Excel.

Random number table

39634	62349
74088	65564
69459	39153
16379	19713
24537	17986

Choose a pattern and stick to it! e.g. every second column of each row

Discard a number if it comes up twice or the number is larger than the population number.

Advantages

- Simple

Disadvantages

- Not the most statistically efficient sampling process
- Possible bad representation of subgroups due to luck

Systematic sampling

- Randomly order your population units and select every n^{th} unit

Procedure

1. Select your population and sampling frame
2. Give each unit in your population a unique random number and sort on this random number
3. Write down your sample size
4. Calculate your sampling fraction – n/N

   N is the total population

   n is the study population
5. Start at a random number within the list
6. Pick every f^{th} value i.e. if your sampling fraction is 1/5, pick every 5^{th} value

Advantages

- Easy and efficient

Disadvantages

- Vulnerable to periodicity

Stratified Random sampling

- Organise your population into distinct categories e.g. male and female
- Randomly sample within each strata using the sampling fraction (or disproportionate sampling methods)
    - you might want to over sample small groups of population to ensure you can get a decent estimate
- Ensures particular groups within the population are adequately represented

Definitions

n1 is the number in the population in category 1

N is the number in the population

f1 is the sampling fraction for category 1: n1/N

Procedure

1. Select your population and sampling frame
2. Choose categories of your population to sample within
3. Assess whether you need over sampling from small categories
4. Calculate your sampling fraction for each category– f = n/N (if no oversampling)

5. Sample n1*f1 from category 1 using simple random sampling
6. Repeat for each category

Advantages

- More efficient and more precision
- Represents key strata proportionately

Disadvantages

- More complicated
    - Have to judge whether small groups should be over sampled and then weighted

Cluster sampling

- If your population is disbursed over a large area you may want to reduce mileage by cluster analysis
- Split your population or geographical region into areas and select a number p of them.
- Each unit of population with these p regions should be measured

Procedure

1. Select your population and sampling frame
2. Split your population into small regions
3. Write down your sample size p
4. Sample p regions using simple random sampling
5. Measure every unit within each sampled region

Advantages

- More economical

Disadvantages

- May not equally represent all responses
- May lose variability information

Multi-stage sampling

- Combine the sampling strategies above to form a multi-stage sampling process
- Similar to cluster sampling, but instead of sampling everyone within a chosen sample you sample within each chosen sample
- For example, sample *n* wards in England (simple random sample), sample *m* streets in those wards (simple random sample) and sample every p^{th} house in each street (systematic sampling)

Procedure

1. Select your population and sampling frame
2. Split your population into small regions
3. Sample regions
4. Sample areas within regions
5. Sample subsets of areas within regions choosing the most appropriate sampling strategy each time.

Advantages

- convenience, economy and efficiency
    - does not require a complete list of members in the target population, which greatly reduces sample preparation cost

Disadvantages

- lower accuracy due to higher sampling error

Quota sampling

- Generally used for street surveys, sample selection is made by an interviewer
- A quota is specified from a subset or set of subsets of the population
- For example, an interviewer might be asked to survey 45 women and 65 men. Once they have interviewed 45 women they will no longer interview women, even if suitable women arrive because the quota has already been filled.

Procedure

1. Select your population and sampling frame
2. Choose your subsets of population
3. Specify a quota for each subset
4. Interview or measure units until the quota is filled for a subset and continue until all other quota are complete.

Advantages

- Quick and cheap to organise

Disadvantages

- Not necessarily random
- Possible biased samples as not everyone gets an equal chance of being sampled

3. Sample size

When deciding on the sample size, there are several issues which need to be considered:

- The objective of the study (e.g. compare types of crossing)
- Identify the null and alternative hypotheses which will meet the objective
- Decide what outcome data are appropriate (e.g. casualty counts)
- Have some idea of the size of effect which would be of interest to detect (e.g. a 5% reduction)
- Have some idea of the variability in the data to be used (from some previous experiments or from a pilot experiment)
- State the precision with which you would want to detect any differences. That is, state the size of the Type I error (probability of rejecting the null-hypothesis when it is true, usually set to 5%), and Type II error (probability of not rejecting the null-hypothesis when it is false, often set to 10%).

Type I and Type II errors

The power of a study refers to the probability that, where there is actually an effect, that this effect will be statistically significant, i.e. the study will conclude that the 'intervention', or whatever is being investigated, had a real effect. It is never possible to be 100% sure that the conclusion reached is correct or not.

There are four possible conclusions when conducting a significance test, which are summarized in the table below.

H_0 = Null Hypothesis of no effect

H_1 = Alternative hypothesis that there is an effect

	Decision	
Reality	Accept H_0	Reject H_0
H_0 is true	OK	Type I error = 5%
H_1 is true	Type II error = 10%	OK

We say that

a. Type I error is made if we reject H_0 when it is true – the probability of rejecting the null-hypothesis when it is true, often set to 5%.

b. Type II error is made if we accept H_0 when it is false – the probability of accepting the null-hypothesis when the alternative hypothesis is true, often set to 10%.

The Type I and Type II errors are often referred to as α (alpha) and β (beta) respectively. Power is defined as (1- β), which is the probability of accepting the alternative hypothesis when it is actually true.

Deciding on a Type II error of 10% suggests that a real effect will be detected with 90% probability – and requires a larger sample than using a Type II error of 50%.

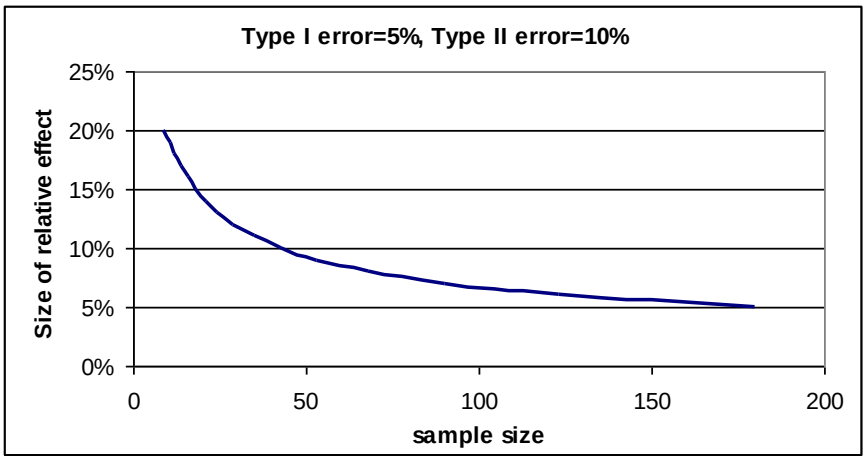
Calculating the sample size

Given the size of effect required and an estimate of the variability within the outcome data, plus the Type I and Type II errors to use – the required sample size can be calculated.

The size of effect required is often not easy to decide upon, and so a curve showing the sample size by the effect size may be helpful in deciding what to use – especially since there are often budget or other constraints on the size of study that can be conducted. If it is not possible to achieve the sample size necessary to achieve the specified accuracy, consider if it is worth doing this experiment and how it could be changed in order to answer the required objectives.

The following graph, Figure 1, is an example using an estimate of the variance from a previous study of pedestrian collisions on crossings.

Figure 1 Size of relative effect by sample



Remember: power of a test is the probability of detecting a difference that we are interested in. Typically a power of 80 or 90% is used.

Power is increased by increasing the sample size.

Sample size (n) depends on the following:

- Size of minimum difference we want to detect, d
- Variability of observations, σ^2
- Required power
- Resources available (time, effort, money) for collecting data

Power and sample size can be calculated by a statistician using modelling techniques but as a rough estimate

$$n \geq \frac{2\sigma^2}{d^2} \times 3^2$$

4. Data sources

There are two types of data source:

Primary data – Collected for the purpose of the experiment so we fully understand any drawbacks it has but often there are not many observations.

Secondary data – Collected for a different purpose. We did not collect the data so it may not really answer the question that interests us. However there may be a very large number of observations.

There are many rich secondary data sources in the UK and a few examples are described below. Primary data collection is discussed in Section 5.

Accident data collection in Great Britain

Since 1949, police throughout Great Britain have recorded details of road accidents that involve personal injury using a single reporting system that is reviewed and updated regularly. The details of the road accidents are stored in a national STATS19 database.

The system records the basic details of the people, vehicles and roads involved in these accidents. A number of variables are recorded for each collision, including details of the collision circumstances, vehicles involved and the resulting casualties (details of what information is recorded is given in STATS20, DfT, 2004). Driver/rider and passenger casualties are linked to the vehicle they were in/on at the time of the collision

Since 2005, contributory factors have been collected which provide some insight into 'why' and 'how' the accident occurred. They are reported by the police officer and represent their view of the key factors leading to the collision. Whilst they give a good indication of the causal factors, they should be treated with caution, many accounts given to the police rely on a subjective assessment rather than hard facts and they are inevitably determined after the accident occurred. There are 77 different factors, and each collision can have up to six factors attributed to it. Contributory factors can be either related to the road environment, behaviour or actions of a driver, rider or pedestrian.

Very few, if any, fatal accidents do not become known to the police. However, research conducted on behalf of the Department for Transport in the 1990s (Simpson, 1996) showed that a significant proportion of non-fatal injury accidents are not reported to the police. In addition some casualties reported to the police are not recorded and the severity of injury tends to be underestimated. The Department is undertaking further

research to investigate whether the levels of reporting have changed. The most recent work by the Department for Transport on levels of reporting and references to earlier reports can be found in articles in two reports (DfT, 2007 and DfT, 2008).

Exposure data collection

A central aim of road safety analysis is to measure and compare the risk of having an accident, so measures of exposure to risk are needed to provide the context for the accident and casualty data. Risk indicators are generally calculated as the ratios between accident or casualty counts and an appropriate exposure measure.

Various measures are used to estimate the exposure to risk of those travelling by road in a country.

- The volume of travel on a country's roads affects the number of road accident casualties, but unfortunately few countries have good statistics about the volume of travel. The number of vehicles in the national fleet generally provides a substitute measure, and it is possible to calculate the traffic volume from the number of vehicles and estimates of the annual average distance travelled per vehicle. Accident risk varies with type of vehicle, being especially high for powered two-wheelers, so information by vehicle type is valuable.

- The characteristics of a country's population, such as the number and age of its residents, directly affect the number of casualties.

- The nature of a country's road network will affect the number of casualties, so if two countries are otherwise similar then the one with the better designed roads will tend to have the fewer casualties. Motorways tend to have fewer accidents than other roads, relative to the volume of traffic, but the high traffic volumes on motorways can mean that they have relatively many accidents per kilometre of road. Vehicle speeds tend to be higher on rural roads than on urban roads, causing accidents to be more serious, so the degree of urbanisation in a country can influence the national casualty data.

Example - UK exposure data

The National Travel Survey (NTS) is the primary source of data on personal travel patterns in Great Britain. The NTS is an established household survey which has been running continuously since 1988. It is designed to monitor long-term trends in personal travel and to inform the development of policy.

The survey collects information on how, why, when and where people travel as well as factors which affect personal travel such as car availability, driving licence holding and access to key services. Data collection consists of a face-to-face interview and a one week self-completed written travel diary. Approximately 20,000 individuals, in 8,000

households, participate in the NTS each year.

A second source of exposure data in the UK is the **National Road Traffic Survey (NRTS)** which provides an annual estimate of the traffic (in million vehicle kilometres). The road traffic estimates are calculated by combining data collected by some 180 Automatic Traffic Counters (ATCs) and 12-hour manual counts at approximately ten thousand sites per annum with road lengths. The figures that relate to traffic are measured in terms of vehicle kilometres. Pedal cycle estimates need to be treated with caution however as they are based on small sample sizes.

5. Data collection

This section presents the advantages and disadvantages of various data collection methods. The data collection system should be well defined, representative of the population of interest and tested.

Data collection by questionnaires or surveys

The most common method for collecting social science data is by means of a questionnaire survey. This consists of a series of questions concerning the facts of an event, the person's life and experiences or their opinions on a subject. The recipient of the questionnaire is referred to as the respondent.

There are four principal methods of collecting questionnaire data:

1. Face-to-face interview
2. By phone
3. By post
4. On-line

Face-to Face interviews

The interviewer and respondent communicate directly, either in a street interview where the interviewer selects passers-by or in an interview in a respondent's home.

Advantages

- The structure of the questionnaire can be **complex** (i.e. if answer is A go to question 6 else question 8)
- Questionnaires will be **completed consistently** as the interviewer completes them
- The interviewer is able to explain the question
- The number of people responding (response rate) is likely to be higher

Disadvantages

- **Expensive** as you have to pay the survey enumerators

- May contain **bias** as the interviewer may consistently misinterpret the question or the interviewer may influence the answer some how
- A respondent may not answer truthfully through embarrassment (e.g. a questionnaire on alcohol or drug use)

The telephone interview

Telephone interviews are occasionally used as a cheaper alternative to face-to-face interviews.

Advantages

- Cheaper than face-to-face interviews

Disadvantages

- Unrepresentative sample – some people may not be in when the call is made (time of day), may put the phone down
- Incomplete sampling frame – around 25% of UK households are ex-directory and as a result do not appear in the sampling frame. The characteristics of the 'missing' people may be different from those on the sampling frame.

The postal questionnaire

The questionnaire is given out for self completion and is returned anonymously by the respondents.

Advantages

- Cheap method of collecting data (no interviewer needed)
- A respondent may be more likely to answer truthfully as anonymous (e.g. a questionnaire on alcohol or drug use)

Disadvantages

- Low response rate often under 50%.
- Possible bias – the replies are unlikely to be representative of the population. People who take the trouble to complete a questionnaire are not typical. The replies can be weighted to allow for this.

Example – A survey to study the accident risk of motorcyclists

A survey was carried out to relate variables, such as attitude, personal characteristics, self-reported riding behaviour and level of experience and training, to accident risk. A single random sample of 30,000 motorcyclists was obtained using vehicle licensing records and a postal questionnaire was sent to them. Reminders were sent a month later and the survey was closed two months after that. The response rate was 40%. Originally the study was intended to collect injury only accidents but as the accidents were self reported many of them were vehicle damage only accidents. To collect injury only accidents the vehicle licensing records

could have been linked to the STATS19 records. However, the names and addresses of riders with injury records are confidential and could not be obtained.

On-line and email questionnaires

The internet and email make it faster and cheaper to contact respondents.

Advantages

- Cheap and fast method of collecting data
- Data accuracy – the responses can automatically be inserted into databases and spreadsheets – reducing data entry errors. Automatic validation checks could be set up to check for errors while the respondent is completing the survey.
- A respondent may be more likely to answer truthfully as anonymous.
- Flexible design – able to direct the respondent to particular questions and the ability to make the survey more user-friendly and attractive.
- Respondents may be more likely to complete an electronic questionnaire and return by email rather than posting a form back.
- A good way of targeting respondents if it is linked to a particular website. E.g. keen cyclists may look at particular cyclist interest sites.

Disadvantages

- Sample Bias – some people do not have access to the internet or do not know how to use it. Possibly a good way of contacting young people but maybe not such a good way of contacting older people.
- Less control over the sample population - there may be several respondents at one computer address or one respondent may complete a questionnaire from a variety of computers.
- Non-response bias - respondents who answer an online questionnaire have very different attitudes or demographic characteristics to those who do not respond.

6. Questionnaire Design

When designing a questionnaire the following points are important.

- Use simple questions
- Keep the length of the questionnaire short
- Start with easy questions
- Ask general questions first and specific questions afterwards
- Open questions are difficult to analyse so consider whether they could be converted into closed questions with multiple answers.

- Ensure all the possibilities are included in a closed multiple choice question
- Do not include unanswerable questions
- Do not include a biased set of multiple choice answers
  e.g. Do you think statistics is
    (a) Very interesting
    (b) Quite interesting
- Do not ask leading questions
- Avoid unfamiliar words
- Avoid long and complicated questions
- Avoid questions about the distant past
- **Ensure that the questions are disaggregated by gender and age**

Checklist Interviews

Undertaking preliminary checklist interviews with key informants can be very helpful for sorting out important questions to ask in a subsequent survey which will ultimately save time and money.

A checklist interview is a good way of identifying key issues that should be included in the survey, and to focus on the issues that respondents really care about.

Checklist interviews offer an opportunity to introduce a new project and its goals and objectives to local residents, to answer their questions, and collect information for triangulation with other data sources.

Pilot Surveys

Any questionnaire used in a project should have been used before or piloted before use. This will help to avoid the use of ambiguous questions or closed questions with missing possible answers. The pilot study uses the entire questionnaire with a small number of people. If the questionnaire is to be completed by interviewers then they should be trained by whoever designed the surveys so that there is consistency in data collection and the responses are provided correctly and uniformly. An allowance should be made for refusals or non-return of questionnaires.

It is vital that every questionnaire is numbered and includes the date and location of the survey, as well as the name of the interviewer and interviewee – this will aid data entry and enable validation of the data with the interviewer/interviewee if required.

Data collection for before and after studies

The same data collection technique should be used for all data measurement, for before/after studies and for all experimental treatments to ascertain pre- and post- intervention impacts (benefits and disbenefits).

Types of Questionnaire Survey

In conducting research relating to transport service provision, there are typically four types of questionnaire survey:

1. Key informant questionnaire
2. Household questionnaire
3. Transport operator questionnaire
4. Travel diary

The following provides guidance on the contents of each type of questionnaire, and the accompanying handout provides a sample of each type of questionnaire. These questionnaires are fairly comprehensive, but can be tailored to suit the specific needs of the research project.

The fieldwork exercise will outline guidance on collecting survey data in the field.

Key Informant Questionnaire:

The key informant questionnaire is also known as a 'village level questionnaire' since the aim of it is to obtain generic information about the community or settlement being studied. Information can be collected on:

- Village size and demography
- Accessibility to essential facilities and extension services
- Availability of transport into and out of the village
- Farming and other types of economic activity

The emphasis of a key informant survey may vary depending on who the key informant is the research topic. For example is can focus on access to education, healthcare, markets, outreach, or expand on different types of transport intervention, for instance the variation in road condition and provision of transport services between wet and dry seasons, and road construction and maintenance activities.

It is important to find a key informant who has the most knowledge (historical and current) about the community and if necessary, access to any documents relating to local population census, as well as farming and economic activity, business enterprises located in the community, and places of worship.

This questionnaire should take no longer than 20-30 minutes to complete, and you should allow the informant to find out any pertinent information and respond to you at a later date, for instance relating to population and the local economy.

Household Questionnaire:

The household questionnaire should be administered to the head of the household or alternatively a senior member of the family. Avoid

interviewing children in the household, unless there are no adults remaining in the household, as they may not have in-depth knowledge of household affairs.

Disaggregate the surveys by gender, age and income to capture demographic variations in mobility.

Information can be collected on:

- Household composition and livelihood
- Household travel patterns
- Trip frequency
- Use of motorised/non-motorised transport

In all surveys a trip is defined as a continuous physical movement from a point of origin to a stopping point – the intended destination. Each trip consists of a one-way journey.

This questionnaire should take no longer than 30-40 minutes to complete. It can be tailored to the specific needs of your research question. For instance you may want to expand the table on transport requirements to include social trip-making categories such as community associations, weddings, funerals, places of worship and leisure activities.

Transport Operator Questionnaire:

The questionnaire is aimed at transport service drivers and operators to elicit information on transport service fares, accessibility issues, vehicle operating costs and transport operator cartels/monopolies where they exist.

Transport service operators and vehicle owners can be interviewed to determine factors affecting their businesses in relation to the routes they travelled on, as well as the quality and standard of roads. The information can be used to determine whether issues relating to cartels, vehicle operating costs, market competitiveness and unnecessary queuing systems affect the frequency and cost of transport service provision. This affects the end user who is unable to transport produce to market, and is subjected to inflated tariffs.

Information can be collected on:

- Vehicle type, operating and maintenance costs, utilisation
- Problems encountered on feeder roads
- Travel patterns in relation to road standard and quality, trip distance and fares

Transport vehicle owners and operators can be interviewed at the villages, along roads leading to the villages, or at selected car and lorry parks.

This questionnaire should take no longer than 20-30 minutes to complete.

Travel Diary:

A travel diary can be administered to a discreet number of households taken from the sample of household questionnaire respondents. Each member of the household completes a logbook of travel every day for a week and records the following:

- Origin/destinations, trip purpose, trip duration, mode of transport
- Household access to vehicles/means of transport
- Income earning trips for one week

This method of capturing travel data is high risk because it relies on the respondent completing the travel diary accurately for every trip made, and remembering the details of each trip. It is recommended that when the travel diary is collected, the respondent describes their journeys to the enumerator who can clarify any anomalies if required.

This questionnaire should be completed whenever travel takes place and as a minimum on a daily basis. Each member of the household should complete a logbook of travel every day for a period of one week.

Ethical Considerations

Sensitisation:

Respondents should be sensitised by informing them of the research being undertaken and how their assistance in the surveys will provide valuable information to decision makers. In order to gain children's consent and involvement in research, approach the appropriate adult 'gatekeepers' (parents, guardians, teachers etc.) who are able to monitor the researchers' access to the children. Where gender sensitivities exist, for instance in Muslim communities, try and ensure that a female researcher interviews women in the household, or arrange women only focus group discussions with a female facilitator. Equally, the same is true for male participants.

Obtaining informed consent:

Obtain consent to carry out surveys from village leaders and permission from respondents to share personal information.

Participant confidentiality and anonymity, personal data:

The privacy of the participant should be respected. The names of individuals featured in the study should be altered. If confidentiality is not considered when reporting, it may allow opinions to be traced to an individual. Ensure that personal data is stored securely.

Manage expectations:

Avoid raising expectations – inform respondents if the research outcomes are unlikely to result in specific interventions for their village, but explain

that it will contribute to a wider body of knowledge with the potential for influencing policy.

Incentives for participation:

In order to compensate survey recipients for their time spent away from productive activities, an incentive (usually in the form of a gift or monetary value if appropriate) can be provided to the head of the villages surveyed, for distribution to participants – this should be factored into the cost of the survey.

Avoid survey fatigue:

Before the field study begins, the researcher should establish that surveys have not previously been undertaken in the same settlements to avoid potential survey fatigue.

7. Weighting the sample

If you have applied a non proportional sampling fraction to some of your categories,

OR

If the data you have collected are not distributed similarly to the population

AND

You want to generalise your results to the population

THEN

You may need to weight your data before analysing.

You can assign small weights to groups that are over-represented and large weights to groups that are under-represented.

Example 1:

From a survey of motorbike riders we have information on whether they have ever had an accident and what type of motorbike they ride.

Accident	Sample
yes	50%
no	50%

It looks as if half the riders have been involved in an accident

Bike	Sample	Population
Sports	50%	80%
Touring	50%	20%

The Touring riders in the sample are highly over-represented compared to the population of bike riders.

Make the estimate from Sports bikes count 4 times as much as that from touring bikes

- In the population Sports bikes are 4 times more common
- Weight Sports by 4 and Touring by 1.

Accident	Sports	Touring
yes	25%	75%
no	75%	25%

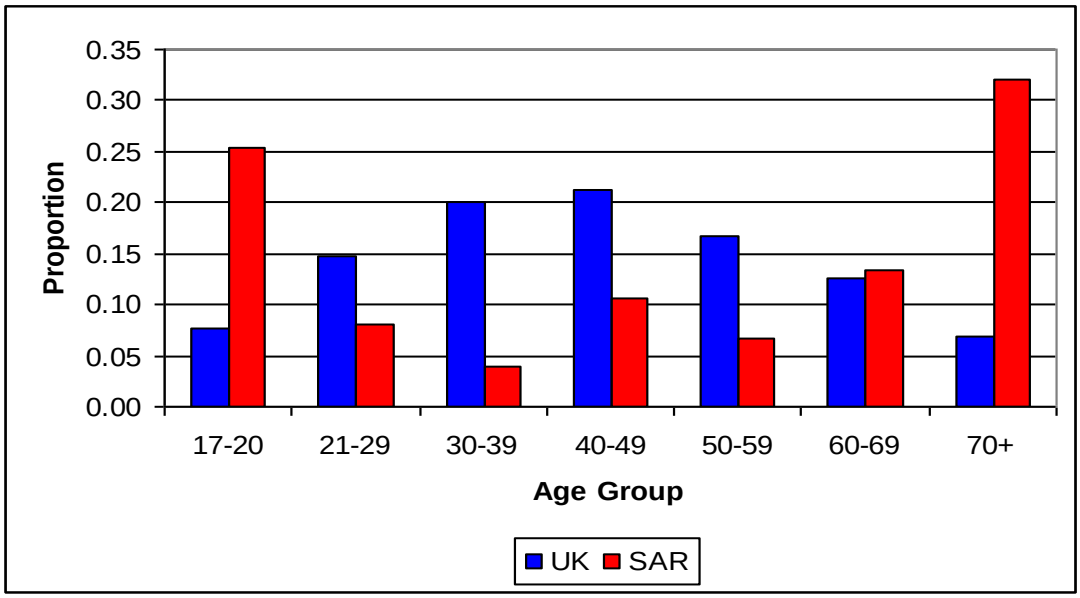
Compute a weighted mean = (25% x 4 + 75% x 1) / 5 = 175 / 5 = 35%

In the previous example it is fairly easy to see that weights would be useful and what the weights should be. Example 2 demonstrates the need to weight the sample so that the population distribution is represented.

Example 2:

Data collected for the static assessment rig (SAR) project were most commonly young and old participants (red).

Distributions of drivers in UK and SAR study by age



We wanted to generalise to the whole population (blue)

Once you have established that the population and sample are not similarly distributed then you may wish to apply some weights. These need to redistribute the sample so that the population distribution is represented.

$$\textbf{Weight} = \frac{\textbf{proportion of population in category}}{\textbf{proportion of sample in category}}$$

Example 2 cont:

Within the age groups the SAR population distributions are "u" shaped, weighting the data allows for this by assigning small weights to those in the extreme age groups and large weights to those in the middle groups.

$$\text{Weight for age group} = \frac{\text{proportion of drivers in UK in age group}}{\text{proportion of people in SAR in age group}}$$

References

Davis, A. 2005. Relationships Between Transport, Mobility, Sustainable Livelihoods and Social Capital for Poverty Reduction. Unpublished PhD thesis, University of Wolverhampton.

Department for Transport (2004). STATS20 Instructions for the Completion of Road Accident Reports. London: Department for Transport

Department for Transport (2007). Road Casualties Great Britain 2006. London: Department for Transport.

Department for Transport (2008). Road Casualties Great Britain 2007. London: Department for Transport.

De Vaus, D. (2002). Surveys in Social Research (5th Edition). London: Routledge

Laws, S. Harper, C. and Marcus, R. (2003). Research for Development: a Practical Guide. London: Sage Publications

Simpson H F (1996). Comparison of Hospital and Police Casualty Data: a National Study. TRL Report TRL173. Crowthorne: TRL Limited

TRL (2004) A Guide to Pro-Poor Transport Appraisal: the Inclusion of Social Benefits in Road Investment Appraisal. Overseas Road Note 22. Crowthorne: TRL Limited

Van Wee, B. (2011). Transport and Ethics: Ethics and the Evaluation of Transport Policies and Projects. Delft: University of Technology