Using Surveys

Quynh Lê, Peter Orpin and Theresa Doherty
Hobart 26th May
Launceston 29th May

Outline

• Making choices
• Types of survey
• Sampling design
• Constructing of a questionnaire
• Analysing data
Making Choices

- Beyond basic sound practice, effective survey data collection involves making a series of decisions/choices:
  - for which there are no ‘pat’ answers;
  - that only you can make, having consideration for your:
    - Resources;
    - Research question and planned information use;
    - Target population;
    - Mode of analysis.

Research Questions

- The hands-on exercises will be based around workshopping one or more research questions.
- You may choose to workshop one your own research questions, or either of the following:
  1. In what ways is risk taking behaviours among year 11-12 females related to their family background?
  2. A study to inform the development of a men’s health education intervention in rural areas

Keep these in mind as you listen!
Types of survey

• There are four main types of survey:
  – Face-to-face administered questionnaire
  – Telephone administered questionnaire (CATI)
  – Mailed or self completed questionnaire
  – Online (Email/Internet) survey

Some questions to consider (SQTC)

  – Which best fits my resources
  – Which is likely to bring the widest and most accurate response from my target population

Sampling Design

10:10 – 11:30 am
Population and sample

- **Population**: is a complete set of people or objects or events which all have at least one characteristic in common, and must be defined specifically and unambiguously.
- **Sample**: any part of the “population” regardless it is representative of the population or not. However, it is important to have a *representative* sample.

Some questions to consider

- What ‘whole group’ of people do I want to be able to generalise my results to – that is, who are the people I want to be able to say something about?
- What are the crucial variables on which my sample needs to be representative for this research – that is, what variations within my sample (vis a vis my population) might stop me being able to apply the results to that population?
Population and sample

Target Population

Sample

All health professionals practising in rural and remote areas of New South Wales

General Practitioners

Some General Practitioners, Nurses, Pharmacists, Dentists, Allied Health professionals, Clinical Psychologists, Social Workers, Specialists etc.

Sampling methods

• **Sampling**: involves decision making on which people (who and how many?) from the population of interest are to be included in the survey.

• **Sampling and Response Bias** happens when some members of the population are less likely to be included than others.

• **Note**
  – bias is most often an unconscious artefact of inadequate methodology – do a risk analysis
  – may only become apparent in the analysis phase
  – where is not feasible to eliminate a source of bias, acknowledge and account for it in analysis and reporting
Sampling/Response Bias - SQTC

• With the chosen methodology, are there significant groups in my population who are:
  – Less likely to receive a survey?
  – Less likely to respond to the survey?
  – Less likely, and/or less able, to provide accurate, complete and/or ‘truthful’ responses?

• How could I modify my methodology to limit any or all of these sources of bias?

• What are my options if a response bias becomes evident in my first-stage analysis?

• What are the implications for my results of any bias I haven’t been able to eliminate?

Sampling methods

• There are two main types of sampling methods:
  – Random: simple random sampling, stratified random sampling, systematic sampling, cluster sampling and multi-stage sampling.
  – Non-random: quota sampling, purposive sampling, convenience sampling, snow ball sampling.
Some questions to consider

• Do I have access to population list from which I can construct a random sample
• Is my research question suited to a random sampling methodology?
  – Am I seeking to generalise from a sample to a population - do I want to use inferential statistical techniques?
  – How is the dependent variable likely to distributed in the population? – rare events

Random sampling

• Simple random sampling
  – Each member of the population has an equal chance of being selected; and
  – The selection of one subject is independent of the selection of any other.
    • E.g. Picking names out of a hat or by assigning everyone a number/code and using a random number table to select the ones for the sample.
  – Simple but not suitable for large population.
Random sampling

• Systematic sampling
  – Individual are chosen systematically but from a randomly chosen starting point in the sample list.
  • E.g. If you want to select 100 names from a population of 1000, you would choose the first name randomly from the first 10 and then every tenth name thereafter.
  – Simple, good spread across the population => know sample size and sampling interval.

Random sampling

• Stratified sampling
  – The target population is divided into groups called ‘strata’. A sample is then drawn from within these strata using simple random or systematic sampling.
  – Some examples of strata commonly used are States, Age, Sex, Religion, Marital Status or Ethnic groups.
  – Better representative of the population but require more information about each group.
Random sampling

• Cluster sampling
  – The target population is divided into groups/clusters. A number of clusters are selected randomly to represent the population, and then all units within selected clusters are included in the sample.
  – E.g. If we are visiting households spread over a large areas, firstly divide the sampling list into areas and then randomly select a sample of these. From this sample of areas, we then randomly choose a selection of individuals in each. The final sample consists of a random sample of clusters of randomly selected individual.
Random sampling

• Multi-stage sampling:
  – like cluster sampling, but involves selecting a sample within each chosen cluster, rather than including all units in the cluster, e.g.
    1. electoral sub-divisions (clusters) are sampled from a city or state;
    2. blocks of houses are selected from within the electoral sub-divisions; and
    3. individual houses are selected from within the selected blocks of houses.

Random sampling

• Multi-stage sampling (cont’d)
  – Convenience, economy and efficiency but lower accuracy due to higher sampling error.

Question to consider:
  – Which sampling methodology will give me the most representative sample within the limits of my resources?
Non-random sampling

• Quota/Structured sampling
  – Individuals are selected according to certain criteria by the researcher to yield a sample which is representative in terms of important basic characteristics.
  – E.g. The researcher might decide to select individuals on the basis of gender and age-groups to reflect the distribution of gender and age-group in the target population. Although these characteristics of the sample may be truly representative, other characteristics may not be.

Non-random sampling

• Purposive Sampling
  – As with quota/structured sample, selection on people or other units are chosen for a particular purpose.
  But
  – Representative through assumed typicality – usually where researching a relatively uncommon characteristic or experience without access to a large sample pool
Non-random sampling

• **Convenience sampling**
  – Anyone who will respond becomes part of the sample.
  – For examples:
    • canvassing opinions prior to designing a questionnaire
    • the first ten people to walk through a turnstile at a sporting event, or
    • females in the first row of a concert.

Non-random sampling

• **Snowball sampling**
  – is achieved by asking a participant to suggest someone else who might be willing or appropriate for the study (*note: SS is subset of purposive sampling*).
  – is used when the desired sample characteristic is rare or from population who individuals are difficult to identify (e.g. victims of domestic violence, refugees, drug users).

**Question to consider:**
• What are the likely sampling and response biases associated with using each of these non-random sampling techniques for your research question?
Sample size (SS)

- Formula
  - \( SS = Z^2 \times (p) \times (1-p) / C^2 \)
  - \( Z \) = Z value (e.g. 1.96 for 95% confidence level)
  - \( p \) = percentage picking a choice expressed as decimal, 0.5 used for sample size needed
  - \( C \) = confidence interval, expressed as decimal (e.g. .04 = +/-4)

- Correction for finite population:
  - new \( SS = SS / (1 + (SS-1)/pop) \)
  - \( pop = population \)
  (Source: http://www.macorr.com/ss_methodology.htm)

Sample size

- Other sources for sample size calculator:

- Garry and Airasian (2003, p193):
  - Population <100, survey the entire population.
  - Population size ~ 500, 50% of the population should be sampled.
  - Population size ~1500, 20% should be sampled.
  - Beyond a certain point (at about 5000 units or more), the population size is almost irrelevant, and a sample size of 400 should be adequate.
Refusals

• Send more survey forms than the estimated sample size in case the returned questionnaires are unexpectedly low.

• Reluctance could be participants’ sensitivity to the questions raised in the survey (e.g. obesity, ethnicity, religion, politics)

Activity 1

• Take either one of the prepared research questions or one generated within your groups and:
  – Choose a survey type
  – Identify the population and set the parameters for a suitable sample
  – Choose a sampling method for the above study.
  – Decide on the appropriate sample size target.

• Be prepared to justify your decisions against the provided list of ‘Questions to Consider’
Construction of a questionnaire

12:00 – 12:30 am
1:00 – 2 pm

Determine the questions to be asked

- What information you need to know;
- What questions should be used;
- Respondents’ ability to answer your questions;
- Respondents’ willingness to answer the questions; and
- Determine in advance how you code the responses.
Independent vs. dependent variables

• There are three main types of information:
  – dependent variables: e.g. people’s attitudes, views, people’s physical health or well-being.
  – independent variables: e.g. people’s attributes (gender, age groups, ethnicity background, education level).
  – confounding variables: Other factors related to both dependent and independent variables that may affect the results and have to be adjusted.

Example: A survey of university students’ attitudes towards risk taking.

– Independent variables: gender, age, ethnic background, family background, academic disciplines.
– Dependent variables: their attitudes towards different risk taking such as smoking, alcohol, sex, gambling, driving.
Some questions to consider

• What variables are of importance in answering my research questions?
• What is my postulated/hypothesised relationship between those variables?
  – Do you think there is any correlation between variables?
  – Do you think there is/are a causal relationship(s) between these variables?
  – What do you think is the direction of any causal relationship?

Simple Rules

• Questions should be clearly and easily answerable!
• Options must be:
  – Fully inclusive – cover all possible responses
  – Mutually exclusive (single choice) – allow only one ‘right’ choice
Question types

• **Closed items**: offers respondents a limited range of specific choices, allowing for reliable data and easy analysis. E.g.
  – Which of the following best describes the benefit you gained from the fitness program that you attended?
    - [ ] Lost weight
    - [ ] Feel better
    - [ ] Improved muscle tone
    - [ ] Made new friends
    - [ ] Improved health conditions

Question types

• **Partial-closed items**: The compromise between open and closed structure. E.g.
  – Please indicate all drugs taken for non-medical used during your pregnancy:
    - [ ] Alcohol
    - [ ] Heroin
    - [ ] Nicotine
    - [ ] Cocaine
    - [ ] Other, please specify:________________
Question types

• **Open-ended items**: allow respondents to answer in their own words. This is ideal for an exploratory study or when the sample size is small. E.g.
  – How could the health services in this community be improved?
  – Any other comments about your health insurance?

• **Scale items**: provide a graded series of responses from which respondents chooses one, e.g.
  – How would you rate your general health (circle one)?
    1 excellent  2 good  3 poor  4 very poor  5 don’t know

• Likert Scale is the most common scaled-response format. It is used to measure attitude or satisfactory.
Question types

• The response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal.
  – Some misuse of Likert scales
    • We have ¼ tsp., ½ tsp., ¾ tsp., or more of satisfaction?
    • The average of fair and good is fair-and-a half? (Jamieson, 2004, p. 1217)

• Can Likert scales be considered interval?

Wording of questions

• Use short and simple sentences: As a rule of thumb, most sentences should contain one or two clauses.
• Ask for only one piece of information at a time (i.e. avoid double barrel question).
• Use their language not yours
• Avoid
  – negative or double negative sentence.
  – ambiguity, confusion, and vagueness: make your questions brief and clear
  – insensitive questions.
  – false premises or presuppositions.
  – unnecessary or unfamiliar abbreviations.
Order of questions

- Be careful not to include so many questions that participants are discouraged to respond.
- Try to motivate participants to complete the questionnaire (e.g. start with fact-based questions and then go to opinion-based questions; use a variety of question formats).
- Group questions into sections.
- Keep open-ended questions to a minimum and position towards the end of the section/questionnaire.

Some questions to consider

- What balance do you need to strike between richness and detail on one hand and ease of entry and analysis on the other?
- Will the questions if answered fully and frankly provide you with the data you need to fully address your research question(s)?
- How would you and yours react when faced with the survey? Take the naïve view:
  - How long does it have to be before you wouldn’t bother?
  - Are you really clear what every question is asking?
  - Does any of it look like duplication on quick reading?
  - Do any of the questions frustrate you because the option provided don’t allow you to answer fully and ‘truthfully’: the ‘Well yes . . and no’ and ‘depends’ answers
Activity 2

• From the given research question in Activity 1, design a questionnaire that include 10 question items which include
  – open-ended questions
  – closed questions
  – scale item questions

• Be prepared to
  – Present your questions for group discussion
  – Justify your decisions against the provided list of ‘Questions to Consider’

Analysing Data

2:15 – 2:45 PM
Processing data

• The main tasks of data processing are coding, entry and checking.
• Some available statistical analysis packages are SPSS, SAS, STRATA, etc.
• Data can be either entered directly or imported from other packages such as Excel or Access.
• Data is held on the computer in a data table (spreadsheet) where each row represents a specific respondent and their data. Each column represents a specific variable.

Data types

• Nominal data are used to reflect different categories of objects, people or entities that are not quantifiable but can only be identified; e.g. hair colour, gender, blood groups.
• Ordinal data are used to reflect a particular order/sequence; e.g. rank in the army, rank in the class.
Data types

• **Interval data** are used to reflect standard and equal units of measurement; However, the zero point is arbitrary; e.g. temperature, calendar years.

• **Ratio data** are similar to interval data and has a true zero point; e.g. income level, distance.

Choosing appropriate statistics

• The way we choose to analyse the data depends on the amount of data; question structure; level of measurement involved; and how you intend to use the information.

• Descriptive analysis describes what the data look like (e.g. where their central midpoint is, how broadly they are spread, how closely the variables within the data are correlated with one another).

• Complex statistical analysis (inferential statistics) allows us to make inferences about large populations by collecting data on a relative small samples.
**Meaning of statistical significance**

- Statistic significance means the statistic is reliable. It doesn't mean the finding is important or that it has any decision-making utility.
- Significance is a statistical term that tells us how sure a difference or relationship exists.
- Significant differences can be large or small depending on your sample size.

**Some questions to consider**

- How are you going to manage the data when you receive it:
  - In what form/using what software?
  - How do you plan to transpose the data into that format?
  - What method/program are you going to use to analyse the data?
- Coding: How amenable are your questions to coding?
- What statistical analytical techniques and reporting forms will you be using?
- Are your data forms amenable to those techniques?
- **Will the data really provide answers to your research question?**
Analysing survey questionnaire using Excel

Demonstration

Introduction

1. Create an Excel database
2. Code your data
3. Enter your data
4. Clean your data
5. Analyse your data
Create an Excel database

• Create a new Excel file
• Create column headers
  – The first column is normally the ID number (should be unique number) of each questionnaire.
  – Create other column headers for each of the survey question items and label them.
• Format the column as appropriate (height, width wrap text, border etc.)

Code your data

• Every response item on the questionnaire needs to be entered as a numbered code (except text from open-ended questions).
• What happens if a question is not answered?
Code your data

2. Do you smoke? (please circle your response.)
   a. No  } 1 = code for “No”
   b. Yes  } 2 = code for “Yes”

3. Do you think second-hand smoke is harmful?
   a. No  } 1 = code for “No”
   b. Unsure } 2 = code for “Unsure”
   C Yes  } 3 = code for “Yes”

   **If unanswered, leave the cell blank.**

4. How many cigarettes do you smoke a day: 25

Enter your data

• What happens if:
  – a question is not answered?
  – the respondent missed out an entire page?
  – two responses are selected for a question when only one is needed?
  – an open-ended question is incomplete?
Clean your data

• Make sure data entry accurate.
• What happens if you have:
  – a small number of questionnaires?
  – a large number of questionnaires?

Analyse your data

• Think about what you want to do with your results:
  – Who will read or use the data?
  – What do they want to know?
  – What types of analysis will they want?
  – What will be the most interest?
  – Will you want charts or graphs to support your findings? (Leady 2004, p. 8)
Frequencies and Percents
- COUNTIF function
- PivotTable Wizard
- Percents: using formulas in Excel

Measures of central tendency: mean, mode and median.
- Mean: use AVERAGE function
- Mode: use MODE function
- Median; use MEDIAN function

Measure of variability (the spread of variation in responses): range, standard deviation, and variance.
- Range (the minimum and maximum answers to a question): Use MIN and MAX functions.
- Standard deviation (the degree to which individual answers vary from the mean): use STDEVP function.
- Variance (has the same purpose as standard deviation) expressed as \((\text{Standard Deviation})^2\): Use VAR function.
Analyse your data

- Cross tabulations: compare the results of two survey questions (Normally, compare independent variable and dependent variable).
- Use PivotTable Wizard to create cross tabulations to calculate frequencies, percents, min, max, standard deviation, variance.
- Charts and figures

<table>
<thead>
<tr>
<th>Goal chart</th>
<th>Measurement level of your variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe a variable or group</td>
<td>Nominal</td>
</tr>
<tr>
<td>Describe a trend</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Present differences between variables or groups</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Analyse association between variables</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Compare observed distribution with normal distribution</td>
<td>Box chart</td>
</tr>
</tbody>
</table>

### Analysing your data

<table>
<thead>
<tr>
<th>Goal analysis</th>
<th>Measurement level of your variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal</td>
</tr>
<tr>
<td>Describe a variable</td>
<td>Frequency table</td>
</tr>
<tr>
<td>Determine the average</td>
<td>Mode</td>
</tr>
<tr>
<td>Analyse distribution</td>
<td>Chi-square test</td>
</tr>
<tr>
<td>Describe groups of cases</td>
<td>Crosstable</td>
</tr>
<tr>
<td>Compare 2 independent groups</td>
<td>Chi-square</td>
</tr>
<tr>
<td>Compare &gt; 2 independent groups</td>
<td>Chi-square</td>
</tr>
</tbody>
</table>


### Analysing your data

<table>
<thead>
<tr>
<th>Goal analysis</th>
<th>Measurement level of your variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal</td>
</tr>
<tr>
<td>Test sample average to another average</td>
<td>Binominal test</td>
</tr>
<tr>
<td>Compare related groups</td>
<td>Wilcoxon Signed Ranks</td>
</tr>
<tr>
<td>Analyse the association between 2 variables</td>
<td>Chi-square</td>
</tr>
<tr>
<td>Explain a variable: Linear relation</td>
<td></td>
</tr>
<tr>
<td>-Non-linear relation</td>
<td></td>
</tr>
</tbody>
</table>

Questions/Answers

15 minutes

References

Thank you

Contact

Quynh.Le@utas.edu.au
Peter.Orpin@utas.edu.au
Bridget.Doherty@utas.edu.au