UNIT 7

DATA PRESENTATION AND ANALYSIS

What this unit is about

This unit provides an introduction to the important concepts used in analysis. It focuses strongly on data, the forms it takes and the ways in which it can be ordered and manipulated to assist in analysis.

Different approaches and techniques are presented that can be used to analyse quantitative data from experiments, physical surveys and questionnaire surveys, and qualitative data obtained through questionnaires, interviews and focus group discussions. Alternative ways of presenting the results of this data analysis are also described.

What you will learn

On completion of this unit, you should be able to:

- list of the different tools and techniques available for data analysis;
- describe various statistical methods that can assist with the analysis of quantitative data, and their costs, benefits and problems;
- illustrate the methods used in qualitative data analysis;
- recognise different types of data presentation styles and be able to present the findings of your data analysis in the most appropriate way.
Contents of this unit

7.1 Introduction ................................................................. 7.4
  7.1.1 Types of data ............................................................ 7.4

7.2 Analysis of quantitative data ........................................ 7.5
  7.2.1 Spreadsheets and databases .................................... 7.5
  7.2.2 Statistics ................................................................. 7.6
    7.2.2.1 Datasets ......................................................... 7.7
    7.2.2.2 Comparing data sets ....................................... 7.8
    7.2.2.3 Sensitivity analysis ......................................... 7.9
    7.2.2.4 Validation ..................................................... 7.9
    7.2.2.5 Errors ......................................................... 7.9

7.3 Presentation of quantitative data .................................. 7.11
  7.3.1 Tables ................................................................. 7.11
  7.3.2 Graphs and charts ............................................... 7.12
  7.3.3 Physical surveys .................................................. 7.13

7.4 Analysing qualitative data ............................................. 7.14
  7.4.1 Analysing interview data ....................................... 7.14
  7.4.2 Data coding .......................................................... 7.15
    7.4.2.1 Numerical data coding .................................... 7.15
    7.4.2.2 Descriptive/analytical data coding .................... 7.16
  7.4.3 What to do after coding ......................................... 7.18
    7.4.3.1 Counting ...................................................... 7.18
    7.4.3.2 Metaphor ..................................................... 7.18
    7.4.3.3 Plausibility ................................................... 7.19

7.5 Presenting qualitative data .......................................... 7.19
  7.5.1 Case examples .................................................... 7.19
  7.5.2 Visual presentation ............................................... 7.20

7.6 Analysing data from maps and tables ............................ 7.24

7.7 Summary ................................................................. 7.24

7.8 References ............................................................. 7.25

7.9 Bibliography ............................................................. 7.26
Figures in this unit

Figure 7.1. Example of a graph ........................................................................................................... 7.8
Figure 7.2. Example of showing error range on a graph................................................................. 7.10
Figure 7.3. Example of a bar chart .................................................................................................. 7.13
Figure 7.4. Example of a preliminary plan for sewerage.............................................................. 7.14
Figure 7.5. Example of descriptive coding....................................................................................... 7.16
Figure 7.6. Example of analytical coding .......................................................................................... 7.16
Figure 7.7. Collection of interview data with tags identifying themes ........................................... 7.17
Figure 7.8. Use of a box to present the summary of a case study ............................................... 7.20
Figure 7.9. Analysis of ranking exercise............................................................................................ 7.21
Figure 7.10. Pie chart ......................................................................................................................... 7.21
Figure 7.11. An organizational chart ............................................................................................... 7.22
Figure 7.12. A community map ........................................................................................................ 7.22
Figure 7.13. A Venn diagram of a community ................................................................................... 7.23
Figure 7.14. Drawing illustrating the layout of a handpump and apron. ......................................... 7.23
Figure 7.15. Photo illustrating quality of latrine construction ....................................................... 7.23
Figure 7.16. An example of a mind map ............................................................................................ 7.24

Boxes in this unit

Box 7.1. Definitions ............................................................................................................................ 7.4
Box 7.2. Statistical definitions .......................................................................................................... 7.7
Box 7.3. What to look for when coding ............................................................................................. 7.18

Tables in this unit

Table 7.1. Slums and poverty in Makassar ....................................................................................... 7.12

Photos in this unit

Photo 7.1 Aerial photograph of Kibera informal settlement, Nairobi, Kenya ................................. 7.25
7.1 Introduction

Without data, analysis is impossible. However, we rarely have perfect data on a given situation and it is often necessary to use judgement when working with the data that have been discovered by research. The first part of this unit provides guidelines for exercising this judgement and then moves on to introduce the different forms in which information might be presented. This is important since analysis must be based on a sound understanding of the type of data available and what might be done with it.

Next, we turn to the assessment of available data – is it reliable and will it tell us what we want to know? The manipulation of this data should give us information and the combination of the information resulting from the analysis should enhance our knowledge, which can then be communicated clearly.

The next part of the unit deals with the organization and manipulation of information. A brief introduction is given to statistical analysis of quantitative data and quantifying qualitative data.

The analysis of data is a crucial part of a research project. Analysis is about organizing data and breaking it down into easily understood parts, which can then be ordered and presented in a form that allows the researcher to answer the initial research questions. The essential component is critical analysis, which involves explanations, comparisons, predictions and exploration of inter-relationships between variables. We will consider how to handle and analyse quantitative and qualitative data. We will also look at different ways of presenting your data and the results of your analysis.

<table>
<thead>
<tr>
<th>Box 7.1. Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data:</strong> recordable facts</td>
</tr>
<tr>
<td><strong>Information:</strong> meaningful combinations of data</td>
</tr>
<tr>
<td><strong>Knowledge:</strong> the sum of what is known by an individual, or about a subject. Knowledge is created through the accumulation of selected items of information. Knowledge is information which has been interpreted and made concrete in the light of the individual’s understanding of the context (World Bank, 1999)</td>
</tr>
<tr>
<td><strong>Communication:</strong> the transmission of data, information or knowledge between two or more points.</td>
</tr>
</tbody>
</table>

Source: Saywell and Cotton (1999)

7.1.1 Types of data

There are four basic categories of data:

- **quantitative data** – which provides details of numbers, sizes or percentages;
- **qualitative data** – which relates to the quality of a service or process, but does not describe it in numerical terms;
- **spatial data** - which is normally presented on maps and plans and provides an indication of where things are;
- **definitive data** – usually provided in the form of a drawing, which defines a particular item and the way in which it has been or is to be built.
The approach taken to analysis will be largely dependent on the form in which data is available.

### 7.2 Analysis of quantitative data

Quantitative data can be a very powerful tool in research. It provides “hard” evidence for a theory that can be proved and tested. Data may be numbers describing such things as:

- Water quantity;
- Population numbers;
- Health status;
- Voting for representatives;
- Water quality;
- Latrine coverage;
- Financial indicators;
- Voting for representatives;

However, the use of supposedly “hard” evidence has to be undertaken carefully. The planning of what data to collect, how to collect it, how accurate it is, how representative it is and how to analyse and present it is a difficult and complex activity. The research must also decide why the data is being collected. Poor use of statistics may give the wrong answer and waste time and money.

*He uses statistics as a drunken man uses lampposts – for support rather than illumination*

Andrew Lang (1844–1912)

Sometimes “proxies” are used as indicators – for example asking people for their job title, when what you want to know is their level of responsibility. Indicators should measure the thing you are interested in as closely as possible. Sometimes numbers can be applied to qualitative data – “Five out of ten people say they prefer VIP latrines to open defecation” for example, or assigning areas of a town as low-, middle- and high-income based on qualitative criteria rather than just household income.

**Further reading:** ‘Indicators’ (DEAT n.d.) in Additional Reading.

### 7.2.1 Spreadsheets and databases

The data will need to be stored, transferred – e.g. typed up from written notes, collated from individual data collection sheets, etc, and analysed. Whilst paper-based methods are acceptable for small amounts of information, computers do offer significant benefits. This allows the data to be copied, stored (keep a copy!) and manipulated quickly and can provide a means of checking the calculations.

Spreadsheets are very flexible and good for small studies, whilst databases can cope more efficiently with larger quantities of data and can be designed to be used by groups of people. Geographical Information Systems (GIS) are a type of database that uses maps to store and display information.

There are also a number of statistical software packages available that make data analysis and presentation much easier. Some (such as SPSS - Statistical Package
for Social Sciences) are specialised. Others are available with standard spreadsheet programmes (for example, the Data Analysis option with MS Excel). However, a fair level of understanding of statistics is needed before using such aids. Other information can be found at websites listed in the bibliography at the end of this unit.

Care needs to be taken with entering data and checking the input is worthwhile. Databases can be designed to check major mistakes as it is entered e.g. only accepting numeric data or one of several options offered in a questionnaire.

Original records of data (e.g. survey notebooks, notes from meetings, questionnaire forms) should be kept even after the data has been transferred, whether to sketch maps, databases, spreadsheets or reports. A weak point in information management is usually the transfer of data from one form to another; keeping the raw data enables checking later.

7.2.2 Statistics

Usually the main purpose of experimental work is to try to demonstrate relationships between different variables (for example, the effect of hydraulic loading rate on the removal of suspended solids from a filter bed). Handling large amounts of data with many different variables is not easy. Careful design of the experimental programme can save considerable time at the analysis stage.

It is important to choose the right form of analysis and data presentation. Both are needed when it comes to providing the evidence you will use to answer your key research questions. Whilst presentation is useful in communicating your findings, using diagrams is also a very useful tool in the analysis itself. Plotting numbers on a graph can show patterns and relationships visually that can then be explored in more detail mathematically. Graphical representations are difficult to communicate other than visually and cannot necessarily be compared. Various mathematical techniques have been developed to try and characterise data, but care has to be taken to ensure the correct statistic is used and that factors such as accuracy, certainty and relevance are correct.

There are three types of lies: lies, damned lies and statistics

Benjamin Disraeli (1804-81)

Another form of data analysis is to construct a model. This can be physical or mathematical. The data obtained from running the model can be used as another set of data and the “real” data compared with the outputs from the model to see if the model does represent reality. The values predicted from the model should match the real data within an acceptable level of error.

Statistical methods are both relevant and useful to the analysis of quantitative data. It is only possible to give a very brief overview of the most common measures and techniques. Most libraries have books devoted to this topic where you can find out more. You can also access a useful guide, Statistical Methods in Water Resources (Helsel and Hirsch, 2002) on the Internet.
### 7.2.2.1 Datasets

Data can be discrete (e.g. total rainfall for February), continuous (e.g. the flow in a river now) or grouped. Grouped data sets can be useful where it is difficult to obtain very precise information or where a high level of precision is not warranted. An example could be an analysis of the time taken to complete contracts for implementing local infrastructure improvements. Here it may be sufficient to set up ranges of say 0 – 3 months; >3 months – 6 months; etc. and state how many contracts fall into each range. A useful finding from this type of simple analysis could be that ‘X per cent of contracts took more than one year to complete’. These groups are called “classes” and form discrete chunks of data. Ideally, the “class size” should be the same – in this case the size is 3 months. Monthly rainfall has an (almost) standard size of 30 days.

### Box 7.2. Statistical definitions

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target population</td>
<td>The whole group under consideration (which needs defining and boundaries setting) e.g. <em>poor rural women</em>.</td>
</tr>
<tr>
<td>Study population</td>
<td>Set of individuals to be studied, e.g. <em>poor rural women in Uganda</em>.</td>
</tr>
<tr>
<td>Population characteristic</td>
<td>The particular aspect we wish to study, e.g. <em>type of water source used</em> by poor rural women in Uganda.</td>
</tr>
<tr>
<td>Cohort</td>
<td>A group of individuals exhibiting characteristics in common</td>
</tr>
<tr>
<td>Sample</td>
<td>A sub-set of the study population that ideally represents the whole study population</td>
</tr>
<tr>
<td>Sampling unit</td>
<td>A basic unit of study, e.g. in an assessment of household water use, sampling unit = household.</td>
</tr>
<tr>
<td>Sampling frame</td>
<td>A list or set of sampling units from which a selection can be made, e.g. electoral register, school attendance list.</td>
</tr>
<tr>
<td>Controlled study</td>
<td>A study of two parallel groups, one with and one without intervention</td>
</tr>
<tr>
<td>Longitudinal study</td>
<td>A study carried out over a number of years that traces change</td>
</tr>
<tr>
<td>Variable</td>
<td>The factor or issue that is under consideration</td>
</tr>
<tr>
<td>Discrete</td>
<td>A variable that can only take specific values (e.g. a hand pump is either working or broken, a road is either concrete, tarmac or earth)</td>
</tr>
<tr>
<td>Continuous</td>
<td>A variable that can be any value within limits (e.g. the well can be 4.58 or 4.61m deep)</td>
</tr>
<tr>
<td>Range or spread</td>
<td>The extreme (maximum and minimum) values of the variable within the population</td>
</tr>
<tr>
<td>Average</td>
<td>An indicator of the value of the variable for the whole population, which can be calculated in several ways</td>
</tr>
<tr>
<td>(arithmetic) mean</td>
<td>The sum total of a variable in the sample (or population) divided by the number in the sample (or population) e.g. total number of children divided by the number of households gives the average (mean) number of children per household</td>
</tr>
<tr>
<td>Mode</td>
<td>The most frequently occurring value (e.g. the blocks of houses in a road had 8, 5, 6, 5, 7, 6, 9, 4, 5, 8, 5 and 4 houses in each, so the mode was 5).</td>
</tr>
<tr>
<td>Median</td>
<td>The middle value of a range (4, 5, 5, 5, 5, 6, 6, 7, 8, 8, 9), 6 is the median</td>
</tr>
<tr>
<td>Variance and Standard deviation</td>
<td>A mathematical calculation to show how wide the spread of a variable is (e.g. two sets of hand dug wells might have the same average depth, but if one set was 4.5, 5.2 and 5.6 and one was 3.3, 5.3 and 7.7 then the second set has a larger standard deviation).</td>
</tr>
</tbody>
</table>
7.2.2.2 Comparing data sets

Quantitative data is often used to compare two (or more) samples. A common research method is to intervene with one sample (e.g. giving soap to a village as part of a hygiene campaign) and comparing it with a “control” sample that did not have this intervention (e.g. still having a hygiene campaign but not supplying soap). Levels of diarrhoea can then be measured to see if there is any impact of handing out soap.

Use of ratios can help to simplify the presentation of data. Let us take the example of the performance of contracts. Suppose we have collected a data set on the costs of the contracts as they were awarded to the contractors, and the final costs that the contractors were actually paid. What we are looking for in our analysis is some measure of the cost overrun. For a particular contract, if we divide the final cost (say $11,000) by the awarded cost (say $10,000) we have a ‘cost overrun ratio’ of 1.1. We can then compare different contracts much more easily than if we simply tried to look at the absolute value of the overrun in dollars. Percentages can also be used in the same way: a 10% cost overrun on a $10,000 project may not be as significant in management terms as a 60% cost over run on a $5,000 project.

Correlation is a way of measuring the association between variables; the measure most commonly used to describe the strength of this association is the correlation coefficient. A correlation coefficient of zero indicates no association; a value of 1 indicates perfect association, a value of -1 shows a negative association (e.g. as the cost of water goes up, the amount used goes down). So in Figure 7.1, there seems to be negative correlation between the time allowed for sediment to settle out of a water sample and the turbidity of that sample.

A common technique for determining the correlation between two variables is simple linear regression. This method fits the best straight line through a data set and uses the correlation coefficient to describe how good the fit is. The turbidity graph could have a line drawn through it to show the relationship; however three lines (0-15 minutes, 15-30 minutes and over 30 minutes) may fit better. A more complex model could use a curve rather than a straight line.

There are more complex techniques, known as multiple regression, which deal with more than two variables.
Further reading: ‘Numerical Analysis’ (Reed 2009) in Additional Reading.

7.2.2.3 Sensitivity analysis

’Sensitivity analysis’ is a useful technique that aims to provide an approximate guide to identifying which variables are the most important in a particular situation. It involves estimating the likely accuracy in particular variables and predicting what effect these will have on any outcome calculated from the variables. Whilst the more obvious application of this method of analysis is to experiments, simple and approximate techniques can be usefully applied to any quantitative data.

For example, suppose we are carrying out a feasibility study for water supply that involves investigating the design and costing of a water distribution network for part of a town. Our approach is likely to involve making a series of important design assumptions about per capita water consumption, population growth, industrial development, design period, pipe materials, cost of finance for capital expenditure etc. We would then use a standard software package to do the actual network design and use these results to estimate the likely capital cost.

Whilst the design may be based on a good working knowledge of the design assumptions, we need to know how important these assumptions are in relation to the final solution, that is, the network design and its capital cost. In this case, the sensitivity analysis would involve carrying out the network design and costing for a higher and lower assumed value of, say, water demand, whilst holding other variables constant. This gives us an estimate of how important the assumptions concerning water demand are in relation to the design and cost of the system. We then do the same for the other assumptions we have made.

What this does is to give us a general idea about which design assumptions have the greatest effect on the outcome, namely on the design and cost of the distribution system. In other words, this method of analysis tells us about the sensitivity of the outcome to specific assumptions. If demand doubled – what impact does this have on the cost? If it does not have a big impact, then do not spend too long getting an accurate picture of demand. This information can be very important in providing guidance about which areas we need to focus on in the feasibility study.

7.2.2.4 Validation

Information can be validated against known data, for example other similar situations or standards, which can give a rough order of magnitude. For example, the number of latrines to be expected in one area might be 5% coverage, so a measured figure of anything between 2 and 15% might be reasonable, but 0.5 or 50% coverage is unlikely, though still possible. This would require checking.

7.2.2.5 Errors

The analysis of experimental error is an important aspect of your data analysis. Part of your experimental methodology involves assessing the degree of accuracy with which you can take measurements. For example, you may be able to measure temperature to ± 0.1 degree Celsius. It is then necessary to investigate how these
errors are compounded in the relatively simple calculations that you carry out on your experimental data. The following general rules apply.

If you add or subtract quantities, the resulting error is the sum of the error in each component; you need to be careful when calculating the percentage errors. For example, with a temperature of 35.2º ± 0.2ºC the percentage error is \((100 \times 0.2/35.2)\) or about 0.6 per cent. For 15.2º ± 0.1ºC, it is about 0.7 per cent. If you subtract the temperatures, the result is 20.0º ± 0.3ºC, giving a percentage error of about 1.5 per cent.

If you multiply or divide different quantities, the resulting maximum percentage error is the sum of the percentage errors in each component. For example, consider the simple formula:

\[
H = \frac{ST}{V}
\]

where S, T and V are variables measured by experiment and the estimated error in each is 2 per cent, 3 per cent and 5 per cent respectively.

The maximum percentage error in \(H = 2+3+5 = 10\) per cent.

Figure 7.2 shows how the magnitude of potential errors can be shown for data points plotted on a graph. The size of the potential error is shown by a line of the appropriate length drawn either side of each data point on the graph. Where the error is only significant in the value measured on the y-axis of the graph, only a vertical line through the point will be used, as on Figure 7.2. If there are also potentially significant errors in the variable plotted on the x-axis, then a horizontal line of the right length will also be drawn through the point to form a cross.

**Figure 7.2. Example of showing error range on a graph**

Further reading: ‘Numerical Analysis’ (Reed 2009: 11-14) the section on Errors, in Additional Reading.

and: various websites providing information on error analysis – see the bibliography at the end of this unit.
Some readings may not conform to the expected pattern and a point may lie outside the expected range of error; this is called an “outlier”. Going back to the original notes may show that a reading was copied down wrong, or it may be a real reading that needs to be examined in more detail and explained.

7.3 Presentation of quantitative data

Whilst a report or a database is useful for recording the data, visual techniques such as graphs, diagrams, maps and tables allow patterns to emerge and provide some sense of order. This is also useful to help identify trends (are death rates going up or down?) and thus help forecast what might be happening. In order to make the data meaningful, it may be useful to provide comparisons, for instance agreed standards or average levels before a disaster. The level of uncertainty should also be communicated.

Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house.

Henri Poincaré (1854-1912)

7.3.1 Tables

This is one of the most straightforward ways of presenting numbers and percentages. This can make cross-comparisons and analysis of data much easier. For example, Table 7.1 provides information on the area of slums, the number of households in those slums and the percentage of those households who are poor, for each District in the city of Makassar. Preliminary analysis of this information reveals that Tallo has the highest area of slums, the second highest percentage slum area and the highest number of slum households.
Table 7.1. Slums and poverty in Makassar

<table>
<thead>
<tr>
<th>District</th>
<th>Area (Ha)</th>
<th>Slum area (Ha)</th>
<th>Ratio: slum to total area</th>
<th>Ratio: poor-HH to total HH</th>
<th>Number of HH in slums</th>
<th>Population in slums</th>
<th>% of poor population in slums</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mariso</td>
<td>182.0</td>
<td>32.40</td>
<td>17.80%</td>
<td>26.51%</td>
<td>2,975</td>
<td>11,901</td>
<td>50.4%</td>
</tr>
<tr>
<td>2. Mamajang</td>
<td>225.0</td>
<td></td>
<td></td>
<td>14.52%</td>
<td></td>
<td></td>
<td>27.45%</td>
</tr>
<tr>
<td>3. Tamalate</td>
<td>2,021.0</td>
<td>56.55</td>
<td>2.80%</td>
<td>8.94%</td>
<td>3,663</td>
<td>14,650</td>
<td>16.33%</td>
</tr>
<tr>
<td>4. Rappocini</td>
<td>923.0</td>
<td>64.72</td>
<td>7.01%</td>
<td>9.37%</td>
<td>3,921</td>
<td>15,684</td>
<td>22.83%</td>
</tr>
<tr>
<td>5. Makassar</td>
<td>252.0</td>
<td>6.25</td>
<td>2.48%</td>
<td>13.75%</td>
<td>258</td>
<td>1,030</td>
<td>49.19%</td>
</tr>
<tr>
<td>6. Ujung Pandang</td>
<td>263.0</td>
<td>5.20</td>
<td>1.98%</td>
<td>3.12%</td>
<td>543</td>
<td>2,170</td>
<td>14.51%</td>
</tr>
<tr>
<td>7. Wajo</td>
<td>199.0</td>
<td></td>
<td></td>
<td>8.45%</td>
<td></td>
<td></td>
<td>40.00%</td>
</tr>
<tr>
<td>8. Bontoala</td>
<td>210.0</td>
<td>23.00</td>
<td>10.95%</td>
<td>9.71%</td>
<td>1,220</td>
<td>4,880</td>
<td>21.43%</td>
</tr>
<tr>
<td>9. Ujung Tanah</td>
<td>594.0</td>
<td>46.81</td>
<td>7.88%</td>
<td>15.56%</td>
<td>2,790</td>
<td>11,160</td>
<td>38.34%</td>
</tr>
<tr>
<td>10. Tallo</td>
<td>583.0</td>
<td>101.48</td>
<td>17.41%</td>
<td>17.90%</td>
<td>7,410</td>
<td>29,638</td>
<td>36.47%</td>
</tr>
<tr>
<td>11. Panakkukang</td>
<td>1,705.0</td>
<td></td>
<td></td>
<td>8.17%</td>
<td></td>
<td></td>
<td>27.42%</td>
</tr>
<tr>
<td>12. Manggala</td>
<td>2,414.0</td>
<td></td>
<td></td>
<td>12.01%</td>
<td></td>
<td></td>
<td>52.94%</td>
</tr>
<tr>
<td>13. Biringkanaya</td>
<td>4,822.0</td>
<td>4.08</td>
<td>0.08%</td>
<td>4.12%</td>
<td>331</td>
<td>1,323</td>
<td>45.42%</td>
</tr>
<tr>
<td>14. Tamalanrea</td>
<td>3,184.0</td>
<td>58.00</td>
<td>1.82%</td>
<td>10.41%</td>
<td>153.75</td>
<td>615</td>
<td>16.33%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17,577.0</strong></td>
<td><strong>398.49</strong></td>
<td><strong>2.27%</strong></td>
<td><strong>11.14%</strong></td>
<td><strong>23,263</strong></td>
<td><strong>33,35%</strong></td>
<td></td>
</tr>
</tbody>
</table>


7.3.2 Graphs and charts

Collecting data is not an end in itself. It must be analysed to see what information it can provide. Numerical data is often best presented in the form of graphs, as it is far easier to assimilate this type of information visually, and can give an idea of how the information can be analysed. It helps display relationships and shows many of the statistical parameters visually. However, the correct style of graph needs to be chosen and decisions over scales and axis can distort the message.

Most ‘spreadsheet’ software, such as MS Excel, allows you to present data in a variety of forms. For clarity, you may find that using colours, and printing the relevant pages of your report in colour, aids communication. However, bear in mind that some people may rely on photocopies of your original, which ideally should also show differentiation between the colours, unless labelling makes the differences clear.

Bar charts can be used for data that is collected in discrete (as opposed to continuous) units, to show the comparison of actual values. Examples of data that may be presented in this way include size intervals, sex, chemical parameters, temperature, or rainfall (see Figure 7.3). The bars can be presented either vertically (as shown in the example) or horizontally, e.g. a Gantt chart (see Unit 3, Figure 3.1).
Figure 7.3. Example of a bar chart

Further reading: ‘Presenting Data’ (Reed 2009) in Additional Reading, for further examples of charts and graphs.

and: Chapter 8 of Northedge (2005: 191-223) for some useful advice on working with numbers, and using tables, diagrams and graphs.

7.3.3 Physical surveys
The results of a physical survey are most conveniently presented on a plan of the area; this requires the results of the different surveys (e.g. level and triangulation) to be worked out. Figure 7.4 shows how the data collected from a survey can be analysed and presented on a plan. It may also be appropriate to prepare cross sections and longitudinal sections of features such as drains and streets. An example of such a physical survey would be a study into different ways of improving the drainage system on a low-income housing site; here it would be possible to show almost all the relevant information on plans and sections.

Numerical data, such as rainfall, can also be presented on a map, using lines of equal rainfall (isohyets) to show the distribution of precipitation over an area.
7.4 Analysing qualitative data

Qualitative data analysis is as much an art as a science. Unlike the analysis of numbers, there is no right or wrong way to analyse opinions, choices, descriptions and feelings. The researcher’s experience and judgement are important tools in making sense of qualitative data. It is just as essential as with quantitative data analysis, however, that analysis is shown to be as rigorous as possible, with methods of analysis clearly explained, and evidence provided to support conclusions drawn.

Keep in mind when preparing to do qualitative research that you need to have a good plan for analysis from the outset, as you are likely to end up with more data than you know what to do with, and in a form that you are not familiar with handling. Be clear about the scope of your work and research parameters, to limit the amount of data collected (see Unit 3).

7.4.1 Analysing interview data

The approach to analysis will depend very much on the type of interview used (Unit 4, section 4.5.3). The more structured the format of data collected, the clearer the structure for analysis.
• A structured interview can be analysed question by question, which are likely to include both quantitative and qualitative data. Highly structured questions can be analysed using numerical coding (section 7.4.2.1), whereas questions with open-ended responses will need to be analysed by identifying themes, common phrases and keywords, i.e. a simple form of coding (section 7.4.2.2).

• A semi-structured interview can be analysed according to themes suggested by the pre-prepared guiding issues/questions (section 7.5.3.2), using simple coding to identify and group the responses.

• An unstructured interview presents the most challenges in terms of analysis, so use this type of interview sparingly, if at all! The structure of the analysis will need to emerge from the patterns and themes identified during the coding process (section 7.4.2.2).

7.4.2 Data coding
Coding is the process of examining the data for themes, categories and keywords, and marking identified chunks (words, phrases, sentences, paragraphs) in the text with a code so that they can be retrieved and collated later for comparison and analysis. This can be applied not only to text such as interview transcripts, but also to other formats such as video. Coding can be numerical or descriptive/analytical.

7.4.2.1 Numerical data coding
A numerical value is given to each variable. This can be an effective way of handling large amounts of data.

**Example:** Do you have a household latrine?
- Yes [0]
- No [1]

Other questions produce more detail, such as if respondents are asked to put things in order of priority (Unit 4, section 4.3.5.2). Responses can be coded by applying a numerical rating to each category.

**Example:** How important do you think it is to have a household latrine? Is it:
- very important [5]
- important [4]
- fairly important [3]
- not very important [2]
- not important [1]

**Example:** At least 50 per cent of the members of the Village Water Committee should be women.
- strongly disagree [5]
- disagree [4]
- neutral [3]
- agree [2]
- strongly agree [1]

This produces what is known as ordinal or ranked data, i.e. that indicates the rank order that something holds but gives no indication of the distance between the points on the scale; nor does the number itself have any absolute significance in numerical terms. It can be used to draw out relative priorities that people assign to a range of issues.
Data analysis software such as SPSS is available for use with structured data from questionnaires. *BUT* a considerable amount of work is required to prepare data for entry into SPSS, so it is really only worthwhile for large amounts of data.

### 7.4.2.2 Descriptive/analytical data coding

This is a way of making sense of data in narrative form, usually from interviews and focus group discussions. Once an interview has been typed up or transcribed (See Unit 6, section 6.9.2), the researcher needs to read the text and identify key themes and concepts. The researcher ‘needs to keep the variation in respondents’ answers, but at the same time sort them into meaningful categories that will aid comparison between respondents and the identification of commonalities (Kumar, 1996: 211).

Coding can be done by circling or highlighting words and phrases in the text, and labelling them with a code – either a number or more usually a code word. Codes may be pre-identified – determined by the research hypothesis or questions, or they may be developed during reading, as themes and patterns emerge (open coding).

---

The most difficult people to change sanitation behaviour were *elderly, children who weren’t involved in the programme & v poor HHs*. This was because of traditional concept *“we have been going out for open defecation for decades & nothing happened”*.

**Figure 7.5. Example of descriptive coding.**

*Source: Author (2008)*

---

The most difficult people to change sanitation behaviour were *elderly, children who weren’t involved in the programme & v poor HHs*. This was because of traditional concept *“we have been going out for open defecation for decades & nothing happened”*.

**Figure 7.6. Example of analytical coding**

*Source: Author (2008)*
Codes are initially likely to be descriptive, i.e. they summarise what the text says, but need to subsequently become analytical, i.e. they represent the researcher’s thinking on why what is occurring in the data might be happening. Figure 7.5 illustrates one method of highlighting and summarising key words and concepts in the text. (Note that space has been left in a wide right-hand margin for this purpose).

In Figure 7.6, themes have started to be identified: in this section of text, the theme identified is that of barriers to take-up of sanitation. These barriers have been further sub-divided into ‘non-participation’, ‘lack of information’ (info.) ‘poverty’, and ‘traditional attitudes’ (trad atts.).

Different coloured post-it notes can be cut up and used to tag pages where themes have been identified (Figure 7.7).

Cut and paste these codes into a Master list to allow for easier analysis across interviews/ units of data. As new codes are created they should be applied to the whole dataset, i.e. also to previously coded units of data. It is better to start with too many codes, which can later be merged if need be, than with too few, and subsequently find that subtleties of meaning have been lost.

Hierarchical coding (tree coding)
Codes can also be arranged hierarchically, like a tree, beginning with the main codes, then adding ‘branches’ representing the various sub-codes. The sub-codes might be ‘examples of…’, or ‘contexts for…’ or ‘causes of…’ etc.

<table>
<thead>
<tr>
<th>People –</th>
<th>Barriers to change -</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Elderly</td>
<td>• Age</td>
</tr>
<tr>
<td>• Children</td>
<td>o Elderly</td>
</tr>
<tr>
<td>• The poor</td>
<td>o Children</td>
</tr>
<tr>
<td></td>
<td>• Non-participation</td>
</tr>
<tr>
<td></td>
<td>• Lack of information</td>
</tr>
<tr>
<td></td>
<td>• Poverty</td>
</tr>
<tr>
<td></td>
<td>• Traditional attitudes</td>
</tr>
</tbody>
</table>

Further reading: Chapter 5 of de Negri and Thomas (2003), and: Gibbs and Taylor (2005) for more ideas on coding and organising interview data.
Box 7.3. What to look for when coding

Most typically, when coding, researchers have some codes already in mind and are also looking for other ideas that seem to arise out of the data. When coding in this second, open minded manner, Charmaz suggests you ask the following questions about the data you are coding:

- What is going on?
- What are people doing?
- What is the person saying?
- What do these actions and statements take for granted?
- How do structure and context serve to support, maintain, impede or change these actions and statements?" (Charmaz 2003: 94-95)

Source: Gibbs and Taylor (2005)

Qualitative data analysis software (NUDIST, NVivo) is useful for analysing unstructured information e.g. interviews, field notes, diaries, audio, film material.

7.4.3 What to do after coding

Once coding is complete, a number of tactics are available for qualitative data analysis.

7.4.3.1 Counting

A certain amount of quantitative analysis (i.e. counting!) will be needed to give an idea of the relative frequency of types of response. Where numbers of respondents are small, e.g. up to 20, using percentages is inappropriate. Instead use actual numbers, e.g. 'eight respondents prefer...'; ‘three of the five female respondents said...’ etc. Or use descriptive phrases such as ‘all respondents’, ‘nearly all respondents’ ‘over half’/ 'less than half' (followed by the exact number in brackets). For example: among 15 female heads of household interviewed, almost all (13) would prefer to have their own latrine.

Frequency is not the only way of deciding what weight to give a category of response. A minority that hold a strong opinion may be significant. Who are the minority? Are they all women, or all of a particular caste, or all disabled or elderly or children? It is important therefore to identify not only what was said, but who said it.

An Excel spreadsheet can be used to tabulate responses according to questions or themes, with a column to record information such as sex, age, caste, ethnicity, chronic poor, disabled, etc. This can allow disaggregation of data according to any number of relevant variables.

7.4.3.2 Metaphor

It is sometimes difficult to structure qualitative data either because it is complex, the data does not map onto a frame of reference easily or it is too abstract. This is where metaphors come in. Metaphors can be used to efficiently convey abstract or difficult concepts by drawing analogies between the data you are explaining and everyday ideas or things with similar characteristics. For example, ‘flying toilets’ is the term used in slums to describe the practice of defecating into plastic bags which are then
thrown by the wayside or as far away as possible. This metaphor is apt as it captures a range of ideas in just two words: the squalor of slum life; the unsafe disposal of human waste; the refusal to take responsibility for one’s own waste; and so on. Use metaphor sparingly and only when the imagery aids explanation. Do not use metaphor in an attempt at creativity!

7.4.3.3 **Plausibility**
Rely on your own good judgement to decide the reliability of what your data is telling you. You do this by applying the plausibility test. In doing the test, rely on your own experience and knowledge, asking yourself, “is this explanation plausible?” If you have done a thorough and careful analysis using techniques outlined above and the explanation based on your judgement is plausible, then it probably is.

Further reading: Chapter 14 of Denscombe (2007) for more information on analysis of qualitative data.

7.5 **Presenting qualitative data**
The main format for the presentation of qualitative data is narrative. This narrative needs to be supported using evidence from the raw data, which can be presented as direct quotes, or paraphrased, to support and illustrate themes. Other ways of presentation include boxed case examples, charts, and images, including photos, diagrams, maps and drawings.

7.5.1 **Case examples**
Qualitative information can be presented as a short case history to illustrate or clarify an issue of concern. Editing of the information is crucial, and must be relevant to the problem or issue under investigation. This is not always easy – on the one hand, we want to give as full a story as possible, but on the other we need to be focused and concise if the key points are to emerge clearly. Figure 7.8 shows how a case history format can be used to present findings from qualitative analysis.
Case history: Example of unsupported initiative for sanitation provision

District: Ranigarathoth
City: Vijayawada, India
Family size: 4 adults, 2 children
Income earners: 3 (Husband Rs.50- per day)
Occupation: Labourer/vegetable vendor

Notes:
This family had been previously relocated from an old bustan site to this district. They had built their own home and provided many of their own services with only limited government assistance. They decided to construct their own latrine (outside of the Municipality’s low cost sanitation programme) because they perceived problems with the programme’s toilets, and did not want to wait for a new latrine construction programme before being able to use their own facility.

The family perceived that the key disadvantage with the programme’s toilets was the need for regular pit emptying, so they constructed a (deep) pour-flush single pit. This facility was built at a time when other construction work on the house was on-going, so exact costs were unavailable. However, in conversation with the householders it was estimated that the total cost, including labour, was Rs5000. A small contractor was employed to build the latrine, and the family saved money from their joint incomes to build the facility. For ten years prior to having a household toilet, the family had resorted to open defecation at a point approximately 200 metres distant. The principal catalyst for latrine construction had been the comfort and convenience it would provide for the users.

Key points:
- Poor perceived reputation of government-led latrine programme (quality, operation and maintenance) affected decision to invest in sanitation (leading temporarily to absence of latrine for the household);
- Socio-cultural factors (comfort, convenience) were principal determinants in latrine construction.

Source: Saywell (2000)

7.5.2 Visual presentation

Qualitative data can also be displayed visually in a variety of ways:

- To provide a succinct summary of the narrative. For example, Figure 7.9 illustrates a comparison of findings from a ranking exercise carried out with two groups of street children in Dhaka. The pie chart shows the percentages of a study population that use different types of latrines (Figure 7.10).

- To portray information that may be difficult or lengthy to explain in words. This could include charts (Figure 7.11), maps and plans (Figure 7.12), diagrams (Figure 7.13), drawings and photographs (Figure 7.14 and Figure 7.15). Free software such as Google Sketchup <http://sketchup.google.com> can be used to give drawings a professional look. Even thoughts can be captured in a diagram, using “mind maps” to show how concepts could be connected (Figure 7.16).
As with all the graphs described in Section 7.2.2, boxes, charts and images serve to illustrate the text, not replace it, so *must always be referred to and explained in the text.*

### Figure 7.9: Analysis of ranking exercise

*Source: Rouse & Ali (2001:42)*

<table>
<thead>
<tr>
<th>Mirpur group</th>
<th>Korail group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Enough food</td>
<td>1. Having a job</td>
<td>Most important</td>
</tr>
<tr>
<td>2. Freedom</td>
<td>2. Enough money</td>
<td></td>
</tr>
<tr>
<td>3. A good house</td>
<td>3. Lots of money</td>
<td></td>
</tr>
<tr>
<td>4. Education</td>
<td>4. A good house</td>
<td></td>
</tr>
<tr>
<td>5. Good health</td>
<td>5. Family</td>
<td></td>
</tr>
<tr>
<td>7. Family</td>
<td>7. Enough good food</td>
<td></td>
</tr>
<tr>
<td>8. Good clothes</td>
<td>8. Good clothes</td>
<td></td>
</tr>
<tr>
<td>9. Having a job</td>
<td>9. Good health</td>
<td></td>
</tr>
<tr>
<td>10. Money</td>
<td>10. Education</td>
<td></td>
</tr>
<tr>
<td>12. Being part of a group</td>
<td>12. Recreation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14. Freedom</td>
<td>Least important</td>
</tr>
</tbody>
</table>

### Figure 7.10. Pie chart
Figure 7.11. An organizational chart

Figure 7.12. A community map
Figure 7.13. A Venn diagram of a community

Figure 7.14. Drawing illustrating the layout of a handpump and apron.

Figure 7.15. Photo illustrating quality of latrine construction.
7.6 Analysing data from maps and tables

Both quantitative and qualitative information can be deduced from non-textual sources, including tables, maps, plans, and aerial photographs (e.g. Photo 7.1.).

Activity: Gathering information from maps

Examine Photo 7.1 which shows an aerial photograph of Kibera, an informal settlement of Nairobi, Kenya. Can you tell the boundaries of Kibera and why? What other forms of development are shown on the image?

7.7 Summary

- Analysis and presentation of experimental error is an essential part of your data analysis when you are dealing with detailed measurements made either in the laboratory or in the field.

- When you are planning your data collection, consider which statistical analysis techniques are likely to be appropriate to the information you are collecting.

- Carrying out a sensitivity analysis on the assumptions you have made often yields valuable insights into which variables are key in any particular situation.
Clear presentation of your data is very important in helping the reader to understand what you have found out; think carefully about the most appropriate ways to present your data.

7.8 References


Google Sketchup <http://sketchup.google.com>


7.9 **Bibliography**


**Websites about statistics:**

http://www.practicalstats.com
http://www.reading.ac.uk/ssc/publications/guides.html

**Websites about error analysis:**

http://www.lhup.edu/~Edsimanek/errors.htm
http://teacher.nsrf.rochester.edu/phylabs/AppendixB/AppendixB.html