



## Chapter 3: Managing Your Data

### Questions

1. When you design your data management approach to your research, do give some time to considering how you will record the responses to questions. You need to know how the data will be manipulated during the analysis later. Numbers can be used as codes in two ways. The first is simply to identify a non-numeric category. These **nominal** numbers are unique identifiers – names, in a way – and do not confer any concept of quantity. Because they do not represent quantity, they cannot be used in calculations. The problem is that any software capable of manipulating numbers in any mathematical way (e.g. spreadsheets) can be instructed to carry out such calculations on any data it recognises as numbers. It is up to you to realise that this is a nonsense. Numbers are used rather than text based codes because computer software finds numeric data easier to handle. Examples of the use of **nominal** numbers will include identification of gender (e.g. male = 1, female = 2), and the identification of a particular response in a range of options (e.g. Devon = 1, Cornwall = 2, Somerset = 3).

Another use of numbers as code is in the form of **ordinal** numbers. These give a sense of position of a response within a range of ordered options. They are more descriptive than **nominal** numbers in that they do more than just identify an option. Typical questions that require this kind of number are those that ask respondents to provide order of preferences or place themselves on a scale (e.g. level of agreement, rating of quality). Ordinal measures convey a sense of ‘more’ and ‘less’ on a response. For example, *Strongly agree* is more than *Agree*, but it is not possible to quantify it in the way we can say that 2kg is twice the weight of 1kg. Chapter 5 looks at **indices** and **scales** and the latter is a good example of the need to use **ordinal** numbers in the coding frame. **Ordinal** measures should not normally be used in calculations unless the number of categories is quite large. More information on this is given in chapter 5. Under these circumstances, **ordinal** measures can be regarded as **interval** measures.

**Interval** measures are quantities in the sense that we generally recognise such a concept. For example a 50 year-old person is twice as old as a 25 year-old. **Interval** measures can also have direction (this was not discussed in the chapter). For example, if respondent ‘A’ has a bank balance of £100 and ‘B’ has an overdraft of £100, we know that while the number in both cases is identical, there is something very different about them. Any question in a social science survey that requires a response to be given as a quantity (e.g. age, salary, length of time unemployed, time since release from prison) will be an **interval** measure. Because they are ‘real’ quantities, we can use them in calculations. This makes them very powerful in data analysis (see chapters 8 onwards). A common error in questionnaire design is to create **ordinal** categories for what are essentially **interval** measures. This happens a lot with the question on age, or salary bands. It would be better to ask the respondent to state actual values rather than place themselves within a range of values. By doing this, the response becomes an **ordinal** measure and so weakens its value during data analysis. Of course, there may be sound reasons for doing this such as you may know from experience that respondents are more likely to respond to grouped options. In such cases, a response to that question is better than no response at all.

2. The coding for the responses to the question is straightforward. The options do not imply either order or quantity, so the coding is a **nominal** measure. What is worthy of further discussion is the use of the *Other* code. This is a valuable addition to questions because, with the best will and effort in the world, there are almost always a range of other possible answers that respondents will want to use. They frequently want to express their own preferences rather than the ones imposed upon them. While this can be irritating, it is also a good way of finding additional categories you have missed. This explains why there are 4 options with the fourth option code being 9. Having used up codes 1 to 3, this leaves codes 4 to 8 available for newly identified options.

The question is, are there enough codes left? There are 5 potential new categories, but more than this may be generated as a result of observed patterns in responses. In that case, it may be better to use codes 01, 02, 03, 99 (i.e. two digit codes). This leaves a very large number of codes available (i.e. from 4 to 98).

Incidentally, if an *Other* category is used in a question, there should be a rider asking respondents to describe what *other* is. This is the only way of knowing how to create new categories. It may also be the case that their perception of *other* is actually a version of one of the formal categories and so you can code accordingly. Chapter 6 looks at questionnaire design in more detail.