

2

Preparing a codebook

Before you can enter the information from your questionnaire, interviews or experiment into SPSS, it is necessary to prepare a 'codebook'. This is a summary of the instructions you will use to convert the information obtained from each subject or case into a format that SPSS can understand. The steps involved will be demonstrated in this chapter using a data file that was developed by a group of my graduate diploma students. A copy of the questionnaire, and the codebook that was developed for this questionnaire, can be found in the Appendix. The data file is provided on the website that accompanies this book. The provision of this material allows you to see the whole process, from questionnaire development through to the creation of the final data file ready for analysis. Although I have used a questionnaire to illustrate the steps involved in the development of a codebook, a similar process is also necessary in experimental studies, or when retrieving information from existing records (e.g. hospital medical records).

Preparing the codebook involves deciding (and documenting) how you will go about:

- defining and labelling each of the variables
- assigning numbers to each of the possible responses.

All this information should be recorded in a book or computer file. Keep this somewhere safe; there is nothing worse than coming back to a data file that you haven't used for a while and wondering what the abbreviations and numbers refer to.

In your codebook you should list all of the variables in your questionnaire, the abbreviated variable names that you will use in SPSS and the way in which you will code the responses. In this chapter simplified examples are given to illustrate the various steps. In the first column of Table 2.1 you have the name of the variable (in English, rather than in computer talk). In the second column you write the abbreviated name

12 Getting Started

Table 2.1	Variable	SPSS variable name	Coding instructions
Example of a codebook	Identification number	ID	Number assigned to each survey
	Sex	Sex	1 = Males 2 = Females
	Age	Age	Age in years
	Marital status	Marital	1 = single 2 = steady relationship 3 = married for the first time 4 = remarried 5 = divorced/separated 6 = widowed
	Optimism Scale items 1 to 6	op1 to op6	Enter the number circled from 1 (strongly disagree) to 5 (strongly agree)

for that variable that will appear in SPSS (see conventions below), and in the third column you detail how you will code each of the responses obtained.

Variable names

Each question or item in your questionnaire must have a unique variable name. Some of these names will clearly identify the information (e.g. sex, age). Other questions, such as the items that make up a scale, may be identified using an abbreviation (e.g. op1, op2, op3 is used to identify the items that make up the Optimism Scale).

There are a number of conventions you must follow in assigning names to your variables in SPSS. These are set out in the 'Rules for naming of variables' box. In earlier versions of SPSS (prior to Version 12), you could use only eight characters for your variable names. The later versions of the program allow you longer variable names, but very long names can make the output rather hard to read so keep them as concise as possible.

Rules for naming of variables

Variable names:

- must be unique (i.e. each variable in a data set must have a different name)
- must begin with a letter (not a number)
- cannot include full stops, spaces or symbols (! , ? * ")
- cannot include words used as commands by SPSS (all, ne, eq, to, le, lt, by, or, gt, and, not, ge, with)
- cannot exceed 64 characters.

The first variable in any data set should be ID—that is, a unique number that identifies each case. Before beginning the data entry process, go through and assign a number to each of the questionnaires or data records. Write the number clearly on the front cover. Later, if you find an error in the data set, having the questionnaires or data records numbered allows you to check back and find where the error occurred.

CODING RESPONSES

Each response must be assigned a numerical code before it can be entered into SPSS. Some of the information will already be in this format (e.g. age in years); other variables such as sex will need to be converted to numbers (e.g. 1=males, 2=females). If you have used numbers in your questions to label your responses (see, for example, the education question in Chapter 1), this is relatively straightforward. If not, decide on a convention and stick to it. For example, code the first listed response as 1, the second as 2 and so on across the page.

What is your current marital status? (please tick)			
<input type="checkbox"/> single	<input type="checkbox"/> in a relationship	<input type="checkbox"/> married	<input type="checkbox"/> divorced

To code responses to the question above: if a person ticked *single*, they would be coded as 1; if *in a relationship*, they would be coded 2; if *married*, 3; and if *divorced*, 4.

CODING OPEN-ENDED QUESTIONS

For open-ended questions (where respondents can provide their own answers), coding is slightly more complicated. Take, for example, the question: *What is the major source of stress in your life at the moment?* To code responses to this, you will need to scan through the questionnaires and look for common themes. You might notice a lot of respondents listing their source of stress as related to work, finances, relationships, health or lack of time. In your codebook you list these major groups of responses under the variable name *stress*, and assign a number to each (work=1, spouse/partner=2 and so on). You also need to add another numerical code for responses that did not fall into these listed categories (other=99). When entering the data for each respondent, you compare his/her response with those listed in the codebook and enter the appropriate number into the data set under the variable *stress*.

Once you have drawn up your codebook, you are almost ready to enter your data. First you need to get to know SPSS (Chapter 3), and then you need to set up a data file and enter your data (Chapter 4).