General Guidelines for Creating a Codebook

Each question on your survey represents one or more variables that will need to be created in your database. Before data entry using SPSS can begin, you need to decide on the following information for each of those variables:

|  |  |
| --- | --- |
| Variable name | Up to 64 characters long, cannot start with a number, and embedded spaces are not allowed. Try to create names that are descriptive enough to be recognizable, but generally no longer than 10-15 characters. For multi-item scales, it is often useful to use a common stem followed by a number (e.g., items on the Piper Fatigue Scale might be named PFS1, PFS2, etc.) SPSS is not case sensitive, so upper or lower case letters can be used. |
| Variable label | A longer (up to 255 characters) text label describing the variable content. Any number of spaces and special characters can be used with labels. Using these helps to document what is in your database. Options can be set so that this information is used to label output from analyses. |
| Format | Numeric (for either a continuous variable or numeric codes), string (text), or date. Numeric is the default format in SPSS, so you will need to change it if the variable has a string or date format. |
| Code values | For any variable that is categorical (or ordinal with a small, defined number of levels), choose numeric codes for each alternative. Data entry is often easiest when codes are 1, 2, 3, etc. for alternatives displayed left to right or top to bottom (but the numbers are arbitrary and using 0-4 instead of 1-5 would be equally acceptable). For questions with yes and no as the only alternatives, coding “1” for “yes” and “0” for “no” is a good choice.  Do not recode during data entry. You want the code to reflect the response chosen by the participant. |
| Missing data code(s) | For numeric variables, choose one or more codes to use if a response is missing. -999 is a good choice. You may want multiple codes, each indicating the reason for missingness (e.g., refused, don’t know, not applicable). |
| Value labels | For each code used for a response alternative, use this text label to describe what the code means (e.g., “1” means “male”; “-99” means “refused”). Usually the wording of the alternatives on the survey can be used or modified as labels. |

Note:

A series of alternatives is presented with the instructions to choose all that apply generally should not be treated as a single variable. Instead, treat it as a series of questions, each being answered yes or no. Therefore, a question such as this with 6 alternatives should be entered as 6 dichotomous variables.