

Encyclopedia of Survey Research Methods

Codebook

Contributors: Lisa Carley-Baxter
Editors: Paul J. Lavrakas
Book Title: Encyclopedia of Survey Research Methods
Chapter Title: "Codebook"
Pub. Date: 2008
Access Date: October 17, 2013
Publishing Company: Sage Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412918084
Online ISBN: 9781412963947
DOI: <http://dx.doi.org/10.4135/9781412963947.n69>
Print pages: 101-102

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412963947.n69>

Codebooks are used by survey researchers to serve two main purposes: to provide a guide for coding responses and to serve as documentation of the layout and code definitions of a data file. Data files usually contain one line for each observation, such as a record or person (also called a "respondent"). Each column generally represents a single variable; however, one variable may span several columns. At the most basic level, a codebook describes the layout of the data in the data file and describes what the data codes mean. Codebooks are used to document the values associated with the answer options for a given survey question. Each answer category is given a unique numeric value, and these unique numeric values are then used by researchers in their analysis of the data.

As a guide for coding responses, a codebook details the question-and-answer wording and specifies how each individual answer should be coded. For example, a codebook entry for a question about the respondent's gender might specify that if "female" is chosen, it should be coded as "1," whereas "male" should be coded as "2." Directions may also be given for how to code open-ended answers into broad categories. These values are then used to enter the data the values represent into the data file, either via computer-assisted data entry software or in a spreadsheet.

There are many ways to create a codebook. Simple codebooks are often created from a word processing version of the survey instrument. More complex code-books are created through statistical analysis software, such as SAS or Statistical Package for the Social Sciences (SPSS). Codebooks generated through statistical analysis software will often provide a variable label for each question, describing the content of the question, word and numeric labels for all answer categories, and basic frequencies for each question.

Codebooks can range from a very simple document to a very complex document. A simple code-book will detail each question-and-answer set along with the numeric value assigned to each answer choice, whereas a more complex codebook will also provide information on all associated skip patterns as well as any variables that have been "created" from answers to multiple other questions.

There are seven types of information that a code-book should contain. First, a short description of the study design, including the purpose of the study, the sponsor of the study, the name of the data collection organization, and the specific methodology used including mode of data collection, method of participant recruitment, and the length of the field period. Second, a codebook needs to clearly document all of [p. 101 ↓] the sampling information, including a description of the population, methods used to draw the sample, and any special conditions associated with the sample, such as groups that were oversampled. Third, the codebook needs to present information on the data file, including the number of cases and the record length of each case. Fourth, the data structure needs to be clearly delineated, including information on whether the data are presented in a hierarchical manner or some other manner. Fifth, specific details about the data need to be documented, including, at the very least, the variable names, the column location of each variable, whether the variable is numeric or character (string), and the format of numeric variables. Sixth, the question text and answer categories should be clearly documented along with frequencies of each response option. Finally, if the data have been weighted, a thorough description of the weighting processes should be included.

Major survey research projects conducted for the federal and state government often create electronic versions of codebooks that are accessible through the agencies' Web sites. There are also numerous centers and libraries at universities that provide archives of survey data from research projects along with Web access to electronic codebooks.

Lisa Carley-Baxter

<http://dx.doi.org/10.4135/9781412963947.n69>

See also

Further Readings

Babbie, E. (2006). *The practice of social research* (11th ed.) . Belmont, CA: Wadsworth.