

How do I create a codebook?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I create a codebook?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149143>

A codebook is a document that outlines the definitions and explanations of variables and codes used in a research project or data analysis. It serves as a reference guide for anyone working with the data, ensuring consistency and accuracy in data interpretation. To create a codebook, one must first identify the variables and codes being used, and then provide detailed descriptions and examples for each one. This should be followed by organizing the information in a clear and user-friendly format, such as a table or spreadsheet. The codebook should also be regularly updated to reflect any changes in the data or coding system. By creating a comprehensive and well-organized codebook, researchers can ensure the reliability and validity of their data analysis.

Codebooks

A *codebook* is a document containing information about each of the variables in your dataset, such as:

The name assigned to the variable
What the variable represents (i.e., its label)
How the variable was measured (e.g. nominal, ordinal, scale)
How the variable was actually recorded in the raw data (i.e. numeric, string; how many characters wide it is; how many decimal places it has)
For scale variables: The variable's units of measurement
For categorical variables: If coded numerically, the numeric codes and what they represent

Codebooks can also contain documentation about when and how the data was created. A good codebook allows you to communicate your research data to others clearly and succinctly, and ensures that the data is understood and interpreted properly.

Many codebooks are created manually; however, in SPSS, it's possible to generate a codebook from an existing SPSS data file.

To get the most out of the Codebooks procedure in SPSS, your dataset should already have variable labels and value labels applied *before* you run the Codebooks procedure. If you are not familiar with variable properties, such as labels or measurement levels, or concepts like value labeling of category codes in SPSS, you should read the [Defining Variables tutorial](#) before continuing.

Creating a Codebook from an SPSS Datafile

Simple Codebook

This codebook method prints most of the information found in the Variable View window. It gives the names, labels, measurement levels, widths, formats, and any assigned missing values labels for every variable in the dataset. It also prints a table with the assigned value labels for categorical variables.

You can generate this simple codebook using the point-and-click menus, or using syntax.

Using the Menus

Open the SPSS datafile. Click **File > Display Data File Information > Working File**. The codebook will print to the Output Viewer window.

Using Syntax

```
DISPLAY DICTIONARY.
```

Detailed Codebook

This codebook method includes all of the same information as the simple method, but also includes options for printing summary statistics as well. Unlike the simple method, you can choose which variables are included in the codebook, and you can choose which variable properties are included in the summary. Also unlike the simple method, the summary information for each variable will be printed in its own table.

You can generate this detailed codebook using the Codebooks dialog window, or using syntax.

Note: This procedure was introduced in SPSS version 17 (source: [SPSS v23 Command Syntax Reference](#)). If you are using an older version of SPSS, this command is not available - it will not appear in the menus, and running the syntax will return error messages.

Using the Codebooks Dialog Window

Open the SPSS datafile. Click **Analyze > Reports > Codebook**. In the **Variables** tab: Add the variables you want in the codebook to the **Codebook Variables** box. To include all variables, click inside the Variables box, press Ctrl + A, then click the arrow button. In the **Output** tab: (Optional) Choose what variable and datafile properties you want to be included in the codebook:

Variable information: By default, includes Position, Label, Type, Format, Measurement level, Role, Value labels, Missing values, and Custom attributes. **File information:** None included by default. **Variable display order:** By default, ordered identically to how the variables are ordered in the file. Can also order alphabetically, by file, or by measurement level. **Maximum number of categories:** By default, limits to 200 categories. In the **Statistics** tab: (Optional) Choose what statistics you want in the codebook. By default, counts and percents will be printed for nominal and ordinal variables, and mean, standard deviation, and quartiles will be printed for scale variables. When finished, click **OK**.

Using Syntax

```
CODEBOOK <variables-names-here>
/VARINFO POSITION LABEL TYPE FORMAT MEASURE ROLE VALUELABELS MISSING ATTRIBUTES
/FILEINFO NAME CASECOUNT
/OPTIONS VARORDER=VARLIST SORT=ASCENDING MAXCATS=200
/STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.
```

Note: When listing the variable names in the syntax, the assigned measurement level must be given in brackets after each variable name: for scale, for nominal, for ordinal.

Example: Simple codebook for sample data

To reproduce this example, download the sample SPSS dataset and SPSS syntax file. Run the syntax file on the sample data. This will add all of the appropriate variable labels and value labels for this dataset.

Problem Statement

When sharing your data with others, it's important that your variables are properly documented. This includes having succinct but descriptive labels for your variables, and labels for any numeric codes used for categories.

If you receive a dataset from a collaborator, you can get an overview of its contents by running the Display Dictionary procedure.

Running the Procedure

To generate a simple codebook for the sample data, click **File > Display Data File Information > Working File**.

Output

Syntax

```
DISPLAY DICTIONARY.
```

Tables

The first table is the **Variable Information** table. This table summarizes variable-level information, including:

Variable name Position (i.e., the order of the columns) Variable label Measurement level (nominal, ordinal, scale) Role Column width Alignment Print format Write format

ARABPSYCHOLOGY.COM

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format
ids	1	ID Number	Nominal	Input	8	Right	F5	F5
bday	2	Date of birth	Scale	Input	20	Right	DATE20	DATE20
Rank	3	Class rank	Ordinal	Input	8	Right	F1	F1
Gender	4	Gender	Nominal	Input	8	Right	F1	F1
Athlete	5	Are you an athlete?	Nominal	Input	8	Right	F1	F1
Height	6	Height (inches)	Scale	Input	8	Right	F5.2	F5.2
Weight	7	Weight (pounds)	Scale	Input	8	Right	F6.2	F6.2
Smoking	8	Do you smoke cigarettes?	Nominal	Input	8	Right	F1	F1
Sprint	9	<none>	Scale	Input	8	Right	F5.3	F5.3
MileMinDur	10	Mile run time	Scale	Input	11	Right	TIME11	TIME11
English	11	Score on English placement test	Scale	Input	8	Right	F6.2	F6.2
Reading	12	Score on Reading placement test	Scale	Input	8	Right	F6.2	F6.2
Math	13	Score on Math placement test	Scale	Input	8	Right	F5.2	F5.2
Writing	14	Score on Writing placement test	Scale	Input	8	Right	F5.2	F5.2
State	15	Are you an in-state or out-of-state student?	Nominal	Input	12	Left	A12	A12
LiveOnCampus	16	Do you live on campus?	Nominal	Input	8	Right	F1	F1
HowCommute	17	How do you commute to campus?	Nominal	Input	8	Right	F1	F1
CommuteTime	18	How long does it take you to commute to campus?	Scale	Input	8	Right	F2	F2
LiveOnCampus	16	Do you live on campus?	Nominal	Input	8	Right	F1	F1
HowCommute	17	How do you commute to campus?	Nominal	Input	8	Right	F1	F1
CommuteTime	18	How long does it take you to commute to campus?	Scale	Input	8	Right	F2	F2
SleepTime	19	Hours of sleep per night	Scale	Input	8	Right	F2	F2
StudyTime	20	Hours of study time per week	Scale	Input	8	Right	F2	F2
enrolldate	21	Date of college enrollment	Nominal	Input	22	Left	A20	A20
expgradate	22	Expected date of college graduation	Nominal	Input	22	Left	A20	A20
Major	23	Major	Nominal	Input	50	Left	A58	A58
RankUpperUnder	24	Class Rank (binary)	Nominal	Input	16	Right	F8.2	F8.2

Variables in the working file

The second table is the **Variable Values** table. This table will only appear if you have value labels defined for at least one variable in your dataset; otherwise, it is omitted. This table prints the name of each variable with defined value labels, and lists each code and associated label for that variable.

Variable Values

Value		Label
Rank	1	Freshman
	2	Sophomore
	3	Junior
	4	Senior
Gender	0	Male
	1	Female
Athlete	0	Non-athlete
	1	Athlete
Smoking	0	Nonsmoker
	1	Past smoker
	2	Current smoker
LiveOnCampus	0	Off-campus
	1	On-campus
HowCommute	1	Walk
	2	Bike
	3	Car
	4	Public transit
	5	Other
RankUpperUnder	1.00	Underclassmen
	2.00	Upperclassmen

Qualtrics users: This procedure works well with survey data that you've downloaded from Qualtrics in SPSS format. Use it to check the coding of your multiple choice items!