

Paper 067-2012

Excel Tool for Coding

Ronald R. Palanca

Mathematica Policy Research, Princeton, New Jersey

Abstract

Open-ended coding is the method of assigning numeric values to verbatim questionnaire responses for the purpose of categorizing the responses for analysis. To perform this task, coders go through the list of verbatim questionnaire responses, and then refer to another list for the assigned numeric values to determine which value to assign. This paper presents a cost-effective way to set up a front end for the coders to do the coding by using a SAS[®] dataset. An Excel spreadsheet with the actual question, the verbatim questionnaire responses, and the lookup codes is created in the same worksheet, and each sheet is categorized by question number. There is no need for a predefined template.

Introduction

An open-ended question is a special form of a multiple- or single-coded question to which a verbatim response is expected. An answer to an open-ended question can run from a single word to a number of sentences and each respondent can give varying answers. Because it is not possible to predict the different responses when designing the questionnaire, the responses have to be back-coded or converted into categories for analysis.

Coding can be done in various ways. Coders could go directly into the data, use a database front-end system, or send the list of edits to the SAS programmer.

This paper will illustrate how to create a coding spreadsheet that shows all of the information that a coder would need to code the open-ended responses. In this scenario, there is no need for a predefined template. All that the SAS program would need is a formatted and properly labeled SAS dataset. The SAS program will use the dataset values, formats, and labels to output the question names, open-ended responses, and codes to use in an Excel coding spreadsheet. Each open-ended question is grouped by worksheet. Please see Figure 1.

ID	Q1_4_SPEC	CODE	NEWTEXT
10003961	Argentine		
10004041	Brazilian		
10003880	Brazilian, Spanish		
10003767	Italian, Mexican		
10004203	Mandarin		
10004067	Mandarin, German		
10003945	Mexican		
10003848	Puerto Rican		
10004180	Puerto Rican		
10004106	Xiang		
10004122	Xiang		

CODE	
1	=Spanish
2	=English
3	=Chinese
4	=Other Languages

Figure 1

How does it work?

The spreadsheet is grouped by open-ended verbatim response. In each worksheet, we will show the open-ended values, the actual question number, and all the associated codes to the open-ended question. The challenge here is to map all the format values to the variables associated with the format.

First, we have to get the related code for each open-ended variable. The program uses proc contents to output the list of open-ended variables. We assign a variable counter to this open list of variables. This variable counter will play an important part when we link the related code to each open-ended variable. Using a standard naming convention for open-ended responses is also a very important component of the SAS program. For this purpose, the dataset we will use will have “_spec” as a suffix to the open-ended variable name: for example, Q1_spec. To output only the open-ended variables from the proc contents output file, we search for the “_spec” in the variable name.

```
Proc Contents Data =Indata NoPrint OUT = Indata2(Keep = Name Label);

Data Indata2;
  Set Indata2;
  Where INDEX(Name,"_SPEC") <> 0;
  NUM = _N_;
Run;
```

The program then refers to the session work library and grabs all the variables from the dataset that have formats attached to them. This creates a list of formatted variables.

```

Proc SQL;
  CREATE Table VarswithFmts
  SELECT Name, Format, Label
  From Dictionary.Columns
  Where Libname = "WORK"
  & Memname = UPCASE("Indata")
  & MEMTYPE = "DATA"
  & FORMAT ^= " ";
Quit;

```

This list will then be merged with the “Indata2” dataset. We still need the actual values for the format names, so we merge our current file with the format catalog. To get the catalog, we use the following:

```

Proc Format Library=Work
  Cntlout=Fmtdetail(keep=fmtname start label);
RUN;

```

This then gives us a dataset with the format name, actual codes for the formats, question label, and the “num” variable. Figure 2 shows the current dataset.

FMTNAME	CODE	LBL	NUM
Q1X	1=Spanish	What languages are spoken in your home?	1
Q1X	2=English	What languages are spoken in your home?	1
Q1X	3=Chinese	What languages are spoken in your home?	1
Q1X	4=Other Languages	What languages are spoken in your home?	1
Q2X	1=USA	In what country were you born?	2
Q2X	2=MEXICO	In what country were you born?	2
Q2X	3=JAPAN	In what country were you born?	2
Q2X	4=PHILIPPINES	In what country were you born?	2
Q2X	5=GERMANY	In what country were you born?	2
Q2X	6=PANAMA	In what country were you born?	2
Q2X	7=CUBA	In what country were you born?	2
Q2X	8=CHINA	In what country were you born?	2
Q2X	9=OTHER	In what country were you born?	2
YESNO	0=No	Is respondent receiving care for his/her condition?	3
YESNO	1=Yes	Is respondent receiving care for his/her condition?	3

Figure 2

Using SAS ODS Tagset, we output each open-ended variable to individual worksheets. To do this, we filter the source SAS dataset based on the open-ended variable and do a proc print on the filtered open-ended data. We also output in the same worksheet a subset of Figure 2 based on the “num” variable. We use the same “num” variable to create macro variables to the open-ended variable’s associated labels and question number from the Indata2 dataset and output in the same worksheet.

Conclusion

This paper creates a cost effective tool that would help a coder to categorize open-ended responses. It puts on one screen all the information that the coder will need to code and all that the SAS programmer will need is a formatted and well-labeled SAS dataset.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please feel free to contact the author:

Ronald R. Palanca: RPalanca@mathematica-mpr.com

Mathematica Policy Research
600 Alexander Park
Princeton, NJ 08543
(609) 799-3535

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the United States and other countries. ® indicates a USA registration.

Other brand and product names are trademarks of their respective companies.