

# *numberNotCompleted2* SPSS macro

johann.jacoby@gmail.com, 20190327

This macro recodes marked(0)/not marked(1) answer option level data for items with a number of response options (or varying number of response options) where respondents are instructed to choose only one option (multiple choice items).

## Input

### Original questionnaire

The items for which this macro is intended would look like this in a questionnaire:

Item 1: One of these statements is correct. Mark the correct one.

- ☐ *Breaking Bad* is the best TV show ever made.
- ☐ *Better Call Saul* is the best TV show ever made.
- ☐ *Hell On Wheels* is the best TV show ever made.
- ☐ *The Sopranos* is the best TV show ever made.

Item 2: In which year was the first season of *Breaking Bad* released? Mark the correct one.

- ☐ 2006
- ☐ 2008
- ☐ 2009
- ☐ 2010
- ☐ 2011

## Raw data

In a data set that *numberNotCompleted2* is applied to, there is one variable for each response option in a test. Thus, the data is recorded at the response option level and not the item level. The variables indicate for each response option (regardless of which item a response option belongs to) whether that response option was marked by the respondent or not. There cannot be missing values on these response option variables: either a response option is marked or not marked. This data structure secures consistency in the records of observation: a response option that is not marked will always lead to the same raw data: a value of 0, regardless of whether

- one other response option for the same item has been marked,
- several other response options for the same item have been marked,

- the item following the current one has 0, one or multiple response options marked
- the item preceding the current one has 0, one or multiple response options marked.

In the manual entering of data from paper-and-pencil administered tests and questionnaires, there is thus no need to

- revise originally correctly entered data for preceding items based on finding out that all following items have no response option marked
- revise originally correctly entered data for preceding response options within the same item based on finding out that several following response options for the same item have been marked
- revise originally correctly entered data for preceding response options within the same item based on finding out that none of the following response options for the same item have been marked
- interpret the same non-response to a particular response option as a correct non-response, an incorrect non-response, one of incorrect multiple responses, a non-response indicative of a non-reached item, or a non-response indicative of a skipped or accidentally omitted item.

In sum, this structure of the raw data eliminates the need to interpretatively consider the within item and across item context of a particular option marking behavior.

Typical data looks like this:

ID	I01001	I01002	I01003	I01004	I02001	I02002	I02003	I02004	I02005	...
fr634dg	0	1	0	0	0	0	0	0	1	...
of65erf	0	1	1	0	0	0	0	0	0	...
utdr43f	1	1	0	0	0	1	1	0	0	...
...	...	...	...	...	...	...	...	...	...	...

The variable names indicate

- the Item a response option variable belongs to ("Ixx"), and
- the response option ("Oxx")

In the data example, the respondent with ID fr634dg has marked option 2 ("Better Call Saul") for item 1, and option 5 ("2011") for item 2.

The respondent with ID of65erf has - contrary to instructions but possible factually correct - marked options 2 and 3 for item 1 ("Better Call Saul" and "Hell On Wheels"), and not answered to item 2.

The respondent with ID utdr43f has also - contrary to instructions but possible factually correct - marked options 1 and 2 for item 1 ("Breaking Bad" and "Better Call Saul"), and also contrary to instructions and definitely factually incorrect - given two answers to item 2 ("2008" and "2009").

## Input and running the macro

Every item must have exactly one correct answer. The number of response options may be different for different items and will be indicated in the macro call as such.

The macro (everything between DEFINE and !ENDDEFINE) needs to be run in SPSS syntax once per session and can then be used as many times as needed. Make sure that the data set that the macro is to be applied to is active.

The macro is called as follows:

```
numberNotCompleted2 prefix = (bestshow) varList = (I01001 I01002 I01003  
I01004 I02001 I02002 I02003 I02004 I02005) nsOfOptions = (4 5).
```

## **prefix**

Indicate a scale name here. This scale name will be prefixed to all variables that the macro creates. This names must only contain non-special character letters and numbers, must start with a letter and needs to be enclosed in parentheses after an equal sign that follows the parameter name "prefix".

## **varList**

is an exact list of the response option variables in the order in which they appear in the data set and must be grouped consecutively as they belong to items. The items must also be in the original order as they will be consecutively numbered later. The keyword TO cannot be used to save space in varList, all variable names must be listed verbatim. Variable names must be enclosed in one opening and one closing parentheses, separated by blank spaces and follow an equal sign after the parameter name "varList".

## **nsOfOptions**

This is the specification of which response options belong to the same item. The nsOfOptions must be integer numbers that identify the number of response options per item in the same order as the variable names in the varList parameter.

These numbers must be separated by blank spaces and be enclosed by one opening and one closing parentheses.

The sum of the numbers in nsOfOptions must be equal to the number of variable names in varList. The number of integers indicated in nsOfOptions is the number of items in the data which cannot exceed half of the number of response option variables.

## Results

If run successfully, the macro creates three different types of new variables.

**numResp**

There will be as many of these items as there are items in the original data. i.e., as many as there are numbers in `nsOfOptions`.

```
<prefix>,"_ITEM_",<n>,"_numResp
```

<prefix> will be the prefix for the scale name specified in the macro call, <n> will be the number running number of the item (e.g., the item numbers in the example data above).

These variables will indicate the number of responses given by respondents to the respective items. In the example data above,

- fr634dg has marked 1 option in item 1, and one option in item 2. They will thus receive "1" on bestshow\_ITEM\_1\_numResp, and "1" on bestshow\_ITEM\_2\_numResp.
- of65erf has marked options 2 and 3 for item 1 ("Better Call Saul" and "Hell On Wheels"), and not answered to item 2, and thus will receive "2" on the variable bestshow\_ITEM\_1\_numResp, and "0" on bestshow\_ITEM\_2\_numResp
- utdr43f has marked two options for item 1, given two answers to item 2, and will receive "2" on the variable bestshow\_ITEM\_1\_numResp, and "2" on bestshow\_ITEM\_2\_numResp.

[illegible]

**theR**

There will be as many of these items as there are items in the original data. i.e., as many as there are numbers in nsOfOptions.

```
<prefix>,"_ITEM_",<n>,"_theR
```

<prefix> will be the prefix for the scale name specified in the macro call, <n> will be the number running number of the item (e.g., the item numbers in the example data above).

These variables will indicate the actual responses given by respondents to the respective items. In the example data above,

- fr634dg has marked 1 option in item 1: option 2; and one option in item 2: option 5. They will thus receive "2" on bestshow\_ITEM\_1\_theR, and "5" on bestshow\_ITEM\_2\_theR.
- of65erf has marked options 2 and 3 for item 1, and will therefore receive -88 on the variable bestshow\_ITEM\_1\_theR (a missing value for "multiple response options marked") as no specific answer can be identified. They have not answered item 2, and thus will receive "-99" on the variable bestshow\_ITEM\_1\_theR, a missing value for "no response option marked".
- utdr43f has marked two options for item 1 as well as for item 2. They will therefore receive -88 ("multiple response options marked") for both bestshow\_ITEM\_1\_theR, and bestshow\_ITEM\_2\_theR.

[illegible]

## blankItemsAtEnd

There will be one variable for one test (i.e., one run of the macro) that indicates for how many consecutive items at the end of the test no response option has been marked.

<prefix>\_blankItemsAtEnd

<prefix> will be the prefix for the scale name specified in the macro call.

This variable will be of use if one wishes to interpret the number of items that have been unambiguously completed (or the number of items that have likely not been reached) in a test.

However, it gives **a purely descriptive number**, namely that number of items in the longest sequence of items for which no response option has been marked counted backwards from the end of the test.

The blankItemsAtEnd variable cannot distinguish between

- items at the end of a test that have effectively not been seen by the respondent (because of lack of time, or because of the respondent quitting the test before the end), and
- items that have been seen (i. e. the respondent did neither quit nor fail to reach these items), but were not answered accidentally or because the respondent intentionally hesitated to mark one of the response options.

In the example data, respondents fr634dg and utdr43f will receive a "0" on this variable, i.e.: for these two respondents there was no item at the end of the test - consisting of two items for illustrative purposes here only - that had no response option marked at all. Respondent of65erf receives a "1", since there is an uninterrupted sequence of one item without any response option marked at the end of the test.

[illegible]

# Notes

The ***numberNotCompleted2*** macro does *not* code item responses into correct or incorrect. This will have to be achieved separately, albeit potentially using the data in the theR variables created by ***numberNotCompleted2***.

***numberNotCompleted2*** does not relieve researchers from decisions about how to treat cases where multiple response options within one item have been marked. Whether such cases are ultimately treated as missing ("respondent did not follow instructions"), plainly incorrect ("respondent did not give the one clear correct answer and failed the item"), or some response options are (post hoc) identified as "also half-correct" in addition to the originally conceived one correct answer, and more importantly how these cases are then treated in the estimation and recovery of factor/scale scores, is entirely up to the researcher and cannot be done by ***numberNotCompleted2***.

***numberNotCompleted2*** is ignorant about whether a response item that has not been marked

- was intentionally not marked because the respondent did not think the response option is the correct one
- was intentionally not marked because the respondent chose to not respond to the entire item altogether
- was accidentally not marked (because respondents accidentally skipped this item in a paper-and-pencil administration, or they confused adjacent answer options)
- was not marked because the respondent never reached this item (as central in a test that uses the number of not-reached items as an indicator of speed or ability)

In particular, it treats a response option always the same rather than interpret it either as an incorrect non-response or a missing value depending on responses to other items and other response options. Any and all decisions on how to interpret a respondent's failure to mark a particular response option (as an indication of lack of ability, a missing value because of an accidentally skipped item, a missing value because of a deliberate choice not to answer to a particular item, an indication of the respective item not having been reached, etc.) will have to be made by the researcher in the context of specific principled assumptions, coding rules and substantive analysis decisions.

***numberNotCompleted2*** provides the basics for such a principled analysis and allows to enter data in the least interpretative way, so that the entered data represents exactly and only what can be observed in responses in a maximally descriptive fashion, rather than put undue interpretative burden and strain on those entering the data and effectively interpret behavior during data entry already instead of recording genuine raw data.