Understanding Society User Support - Support #2224

SF12 coding

03/13/2025 09:45 AM - Mark Bryan

Status:	Feedback	**Start date:**	03/13/2025
Priority:	Normal	**% Done:**	50%
Assignee:	Understanding Society User Support Team		
Category:	Data documentation		

Description

Hi,

I am interested in the coding of the SF12 variables, sf12pcs_dv, sf12mcs_dv and their sub-components.

I understand sf12pcs_dv and sf12mcs_dv are both scored positively, i.e. higher values indicate better health, and this appears to be confirmed in the data, e.g. sf12pcs_dv declines as people age.

But according to the questionnaire and Stata value labels, scsf1 (general health) is scored negatively, with values ranging from 1 (excellent) to 5 (poor). However, in the data, scsf1 declines with age, suggesting that it is in fact positively scored. Could you confirm? And could you also confirm the coding of the other sub-components scsf2a-scsf7?

Many thanks
Mark Bryan

History

#1 - 03/13/2025 11:17 AM - Understanding Society User Support Team

- Category set to Data documentation

- Status changed from New to Feedback

- % Done changed from 0 to 50

- Private changed from Yes to No

Hello Mark,

The response options for four items (scsf1, scsf5, scsf6a, scsf6b) are reverse-coded so that higher values reflect better functioning. The scores are then rescaled, standardised, and aggregated into summary scores to produce the Physical Component Summary (PCS) and the Mental Component Summary (MCS) resulting in a continuous scale with a range of 0 (low functioning) to 100 (high functioning).

The code for generating this variable can be found in the UKHLS derived variables syntax section under sf12pcs_dv (https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/main-survey/syntax/stata/stata-sf12-dv-public.do)

I hope this information is helpful.

Best wishes,
Roberto Cavazos
Understanding Society User Support Team

#2 - 03/13/2025 12:41 PM - Mark Bryan

Hello Robert

Thanks for the response and code.

In the code, when scsf1 (for example) is reverse coded, a new variable scsf1r is created and scsf1 is left unchanged. Does this mean that scsf1 in the released data is also unchanged, i.e. higher values reflect worse functioning? As I said previously this is what is implied by the Stata value labels (see below). However, I believe that in the released data, higher values reflect better functioning as you suggest. So does that mean the Stata value labels are wrong?

. ta scsf1

```
General |
      health |      Freq.     Percent        Cum.
-------------+-----------------------------------
     missing |        280        0.05        0.05
```

```
inapplicable |     28,200        5.44        5.50
       proxy |     26,142        5.05       10.55
     refusal |        619        0.12       10.67
  don't know |        286        0.06       10.72
   excellent |     63,843       12.33       23.05
   very good |    157,117       30.33       53.38
        good |    145,189       28.03       81.41
        fair |     70,354       13.58       95.00
        poor |     25,922        5.00      100.00
-------------+-----------------------------------
       Total |    517,952      100.00
```

Thanks
Mark

#3 - 03/14/2025 01:54 AM - Understanding Society User Support Team

Hello Mark,

The released variables follow the structure presented in the survey questionnaire. You can refer to the Self-Completion SF12 Module here: https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaire-modules/scasf12_w14/#scasf12_w14.scsf1 As a result, response option values are not reversed.

For scsf1 – general health, higher response values indicate worsening health. Reverse-coded variables are not publicly released; they are only created internally to compute the physical and mental component summaries (sf12pcs_dv and sf12mcs_dv).

I hope this information is helpful.

Best wishes,
Roberto Cavazos
Understanding Society User Support Team

#4 - 03/14/2025 08:54 AM - Mark Bryan

Thanks for the clarification Robert.

Mark