
Subject: Duplicated label categories

Posted by [ozak](#) on Wed, 05 Jun 2019 20:54:16 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi,

I noticed that e.g. in file ZWIR52DT.ZIP, the included Stata file has duplicated code for some labels. E.g. in variable h46a_1 field worker is coded as 15 & 25 (see below). Any reason for this? Are these different types of field workers or is this an error?

```
. tab h46a_1
```

```
      place first sought |  
      treatment for fever |   Freq.   Percent   Cum.  
-----+-----  
government health center |         3     0.75     0.75  
government health post |        41    10.20    10.95  
  mobile clinic |        75    18.66    29.60  
    field worker |        60    14.93    44.53  
    other public |        17     4.23    48.76  
private hospital, clinic |         1     0.25    49.00  
    pharmacy |        27     6.72    55.72  
  private doctor |         9     2.24    57.96  
private mobile clinic |        49    12.19    70.15  
    field worker |         1     0.25    70.40  
    other private |        81    20.15    90.55  
      shop |        14     3.48    94.03  
      other |        24     5.97   100.00  
-----+-----  
Total |       402   100.00
```

```
. tab h46a_1, nol
```

```
place first |  
  sought |  
treatment |  
for fever |   Freq.   Percent   Cum.  
-----+-----  
12 |         3     0.75     0.75  
13 |        41    10.20    10.95  
14 |        75    18.66    29.60  
15 |        60    14.93    44.53  
16 |        17     4.23    48.76  
21 |         1     0.25    49.00  
22 |        27     6.72    55.72  
23 |         9     2.24    57.96  
24 |        49    12.19    70.15  
25 |         1     0.25    70.40
```

26	81	20.15	90.55
31	14	3.48	94.03
96	24	5.97	100.00
-----+-----			
Total	402	100.00	

Subject: Re: Duplicated label categories

Posted by [ozak](#) on Wed, 05 Jun 2019 21:40:56 GMT

[View Forum Message](#) <> [Reply to Message](#)

Seems this issue is more general and happens in many datasets, which makes importing into programs for preprocessing (e.g., Python Pandas) problematic.

Subject: Re: Duplicated label categories

Posted by [Bridgette-DHS](#) on Wed, 26 Jun 2019 17:17:05 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Stata Specialist, Tom Pullum:

For this variable, first digit "1" indicates "public" and first digit "2" indicates "private". Code 15 is used for public fieldworkers and code 25 for private fieldworkers.

Subject: Re: Duplicated label categories

Posted by [ozak](#) on Wed, 26 Jun 2019 20:34:50 GMT

[View Forum Message](#) <> [Reply to Message](#)

Thank you for the reply. I think it would be useful if this was included in the label, which would help researchers in their interpretation and also prevent errors importing in Python or other software.

Subject: Re: Duplicated label categories

Posted by [ozak](#) on Mon, 05 Aug 2019 16:19:48 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following up on your answer...Do you know in other cases what the difference may be? E.g. in file 'AOIR71DT.ZIP' the variable "'v129': Main material of the roof. Individual codes are country-specific, but the major categories are standard" has duplicated labels for "wood" and a strange label for the potential answer 97 NOT DE JURE RESIDENT:

```
{10: 'natural',  
 11: 'no roof',  
 12: 'grass/palm',  
 20: 'rudimentary',  
 21: 'palm/bamboo',  
 22: 'wood',  
 23: 'cardboard',  
 30: 'finished',  
 31: 'zinc plates',  
 32: 'wood',  
 33: 'calamine/cement fiber',  
 34: 'ceramic tiles',  
 35: 'concrete slab',  
 36: 'tiles',  
 96: 'other',  
 97: 'not a dejure resident'}
```

Why are there two codes for wood? Are these specific types? Or error in the country coding conversion? Where can I find this information? The codebooks seem to only have the variable names, but not the codes and labels.
