Subject: Operationalization v476
Posted by erood on Thu, 01 Sep 2016 09:52:20 GMT
View Forum Message <> Reply to Message

I have a inquiry regarding the recoding of the variable v476 which refers to the question whether a person would keep it a secret if a family member was diagnosed with TB.

I have reproduced the exact rates reported in various DHS reports by recording entries 8 or 9 to a 0. Setting these to missing values therefore results in rates far higher than those reported in the DHS reports.

My question is what was the reason that entries of 8/9 are recoded to 0. This seems to be uninformative resulting in biased estimates. I cannot find any documentation on this matter and would like to know whether I should cohere to this practice when analyzing the DHS data?

Regards, Ente Rood

Subject: Re: Operationalization v476
Posted by Bridgette-DHS on Thu, 01 Sep 2016 15:35:56 GMT
View Forum Message <> Reply to Message

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

First, the code "." means "Not Applicable (NA)" and of course should be omitted from any denominators.

Second, code 9 is not a "valid" response for v476, but it occurs occasionally for many variables for which it is not a valid response, and usually means something like "not stated." and I would agree with you that most analysts would want to omit code 9 and treat it the same as NA. However, for DHS tabulations, it is normal to retain the invalid "9" or "99", etc., in the denominator. I have made my opinion known around here, and lost, but it really doesn't bother me because the number of "9"s is always low.

Third, code 8 or other responses that mean "don't know" or "undecided", or "depends", especially for an attitude question, in my opinion should NOT be removed from the denominator. If you remove those cases, then the balance between "No" and "Yes" can be misleading. But if you keep the "don't know" cases, what do you do with them?

One possibility would be to list them explicitly as a third category. This would be my preference. However, that would cause problems if you wanted to do a logit regression, say. Another possibility would be to divide them evenly between the 0 and 1 categories, but that is a completely ignorant, know-nothing way to divide them. Another possibility would be to divide them between 0 and 1 in proportion to the observed balance between 0 and 1, but that would be equivalent to removing them entirely, i.e. assigning them to NA.

The prevailing practice with DHS would basically be to group the "don't know" cases with the "no" cases. A good example of this is with hiv03, or HIV status, in the AR files. In some surveys you will find a small number of cases in which the HIV blood test was ambiguous or inconclusive. Those cases are not removed, but are classified with the "HIV negative" cases. At one time I disagreed with this practice, but now I'm more accepting of it, because there is an implicit null hypothesis that the person is HIV negative, and if the test result is ambiguous, then it makes more sense to say that it is consistent with the null hypothesis than to ignore it completely. For the question in your example, I think the same reasoning would apply: you should only count a person as a "yes" if they SAY "yes". If they don't say "yes", then count them as a "no" even if they don't quite say "no".

Fortunately, when you do a re-analysis of the data, you have access to the data files and you are free to re-interpret as you wish!