
Subject: Regional variable v024

Posted by [eugenio.cassese@uniboccon](#) on Tue, 09 Apr 2024 11:47:28 GMT

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

I am working with dhs data of several country (the individual surveys) and I often find that the regional variable v024 (and even other variables such as religion) sometimes present values that are not codified. E.g., in Bangladesh the v024 variable is supposed to go from 1 to 6, but tabulating the variable this is the result I get:

region	Freq.	Percent	Cum.
-----+-----			
barishal	12,084	11.23	11.23
chittagong	17,964	16.70	27.93
dhaka	21,921	20.38	48.31
khulna	15,446	14.36	62.67
rajshani	18,963	17.63	80.29
6	12,253	11.39	91.68
7	6,719	6.25	97.93
8	2,229	2.07	100.00
-----+-----			
Total	107,579	100.00	

Are these human errors when compiling datasets or are there perhaps other regions that are not indicated in the dictionary? If so, where can I find such a list?

P.s.: the question extends to all other variables for which this scenario might occur.

Many thanks in advance!

Subject: Re: Regional variable v024

Posted by [Janet-DHS](#) on Wed, 10 Apr 2024 13:03:30 GMT

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

Following is a response from DHS staff member, Tom Pullum:

There have been several previous posts on what happens when you append files from different surveys, even within the same country. I believe the problem is that you have combined (appended) files with different labels for v024. When you do this, the last label will over-write the earlier labels. In your example, it appears that one of the surveys included regions numbered 6, 7, and 8, but the label for the last file in the append DID NOT include those regions. As a result, categories 6, 7, and 8 have numerical codes but no labels.

One way to handle this is to give a new name to v024 in each survey, such as v024_BD7R, and a new name to the label. After you have your final file, you can do a recode which synthesizes the different v024_* variables into a single variable. But sometimes a synthesis is impossible.

Many other variables, such as source of drinking water, types of health facilities for treating child illness, foods, etc., are coded differently in different surveys, and must be recoded. You have to check the actual labels in each survey, for example with "label list V024". Sometimes the codes will shift even if the names of the regions are stable.

Subject: Re: Regional variable v024

Posted by eugenio.cassese@uniboccon on Fri, 12 Apr 2024 16:43:17 GMT

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

Many thanks to the clarification. As a follow-up, is it then possible to have a situation in which 2 variables in different countries/years have the same name but different contents? (e.g., v738 being highest level of education in Bangladesh 1996 and being used contraceptive method in India 2011; I am making up names and contents)

Subject: Re: Regional variable v024

Posted by [Janet-DHS](#) on Fri, 19 Apr 2024 20:58:33 GMT

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

Following is a response from DHS staff member, Tom Pullum:

The v variables (variables with prefix v, hv, or mv) are fixed (so far as I know). They will mean the same thing in every survey. However, the specific codes can be different. Source of drinking water would be an example. It has the same variable name in every survey but the codes and categories can be different. This generalization would apply to almost all variables except the survey-specific variables that begin with "s".

As a general rule, the more recent the surveys, the greater the standardization. A good thing about the IPUMS files is that their staff have checked very carefully for potential inconsistencies and have developed even more consistent labels and codes than in the DHS files.
