
Subject: Re: Education in single years
Posted by [Janet-DHS](#) on Tue, 28 Jun 2022 13:39:44 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Probably the best way to check the interpretation is to go to the questionnaire. In the 2019 household questionnaire, which is included as an appendix to the final report, the question is "What is the highest level of school (NAME) has attended? What is the highest grade or year (NAME) completed at that level?" this is a two-part question that is coded into hv106 and hv107, respectively. Then hv108 is constructed during data processing from the combination of hv106 and hv107.

Here are the labels for these three variables:

hv106 highest educational level attained
hv107 highest year of education completed
hv108 education completed in single years

To see how they are related, that is, how the third variable is constructed from the first two, I do some "reverse engineering" with the following Stata command: "tab hv107 hv106, summarize(hv108) means" , which produces the following table of means of hv108 for each combination of hv106 and hv107:

Means of education completed in single years

highest year of highest educational level education attained completed primary secondary higher Total				
0	0	8	12	1.928264
1	1	.	13	2.3707113
2	2	.	14	3.3890339
3	3	.	15	6.5174477
4	4	.	16	5.5426389
5	5	.	17	5.7106918
6	6	.	18	6.4073853
7	7	.	19	7.1210593
8	8	8	20	8.0169252
9	.	9	21	9.0375783
10	.	10	22	10.023841
11	.	11	23	11.195652
12	.	12	24	12.75
don't kno	98	.	98	98
-----+-----+-----				
Total	3.9182786	9.9825762	15.239357	5.9666494

Ignore the "Total" column and the "Total" row! The table shows that hv108 values 0 through 24 can be interpreted as "highest grade" completed, as the label states. It is not the number of years attended school, which you might have been thinking. The construction is appropriate if "secondary" begins with grade 8. I believe that in some settings, secondary begins with grade 7.

I suspect that the higher values are reporting errors for some cases. Some of them are incompatible with current age. When you work with this variable, understand that there is a code 98, which must be omitted from any calculations such as means.