
Subject: Mismatch in educational attainment and education in single years

Posted by [Navya](#) on Tue, 09 Jan 2024 12:08:24 GMT

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

Dear all,

I am currently working with the men's and individual's recode files for the 2019-21 survey wave for India, and I noticed that there may be some inconsistency between the educational attainment variable (mv/v149) and education in single years variable (mv/v133) for both files. From what I understand, (m)v133 gives the single years of education for the individual, going from 0 years to 20 years for both files, and (m)v149 categorizes these years into "no education", "incomplete primary", "incomplete secondary", "complete secondary", and "higher". According to this, individuals who completed 11 years of education should be categorized into "incomplete secondary" and those with 12 years should be categorized into "complete secondary". However, this result seems to be flipped (individuals with 11 years are being categorized into "complete secondary" and those with 12 years are under "incomplete secondary" for both files). I have attached a screenshot of the tabulate result I get in Stata for the individual's recode file. I only see this result in the 2019-21 survey datasets for India, and not in the older datasets. Am I missing something here?

Thank you for your help.

File Attachments

1) [Screenshot 2024-01-09 173008.png](#), downloaded 290 times

Subject: Re: Mismatch in educational attainment and education in single years

Posted by [Bridgette-DHS](#) on Tue, 09 Jan 2024 14:20:22 GMT

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

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

I'll describe the variables for men in the MR file, with the "mv" prefix. The same applies for women in the IR file, with the "v" prefix.

There are four relevant variables for education:

. describe mv106 mv107 mv133 mv149

	storage	display	value	
variable name	type	format	label	variable label
mv106	byte	%8.0g	MV106	educational level
mv107	byte	%8.0g	MV107	highest year of education (at level in mv106)
mv133	byte	%8.0g	MV133	total number of years of education
mv149	byte	%8.0g	MV149	educational attainment

mv149 and mv107 are almost exactly the same. The only difference is that mv149 divides primary into incomplete and complete:

```
. tab mv149 mv106
```

educational attainment	educational level				
	no education	primary	secondary	higher	Total
-----+-----+-----					
no education	12,269	0	0	0	12,269
incomplete primary	0	11,710	0	0	11,710
incomplete secondary	0	0	55,807	0	55,807
complete secondary	0	0	4,211	0	4,211
higher	0	0	0	17,842	17,842
-----+-----+-----					
Total	12,269	11,710	60,018	17,842	101,839

mv133 is a recode of mv106 and mv107. The following table gives the mean of mv133 for each combination of mv106 and mv107. (Ignore the "Total" column and row.)

```
. tab mv107 mv106, summarize(mv133) means
```

Means of total number of years of education

highest year of education	educational level			
(at level in mv106)	primary	secondary	higher	Total
-----+-----+-----				
no years	0	.	.	0
1	1	6	13	8.5814143
2	2	7	14	8.1771985
3	3	8	15	10.65447
4	4	9	16	8.6599261
5	5	10	17	9.5647399
6	.	11	18	11.566783
7	.	12	19	12.099424
8	.	.	20	20
-----+-----+-----				
Total	3.9836038	9.5509347	15.147237	9.9378475

Hope this answers your question. Please let us know if it does not.

Hi,

Thank you so much for your quick response. I am still having some trouble reconciling the values for (m)v149 (educational attainment) with (m)v107 (highest year of education (at level in (m)v106)) and (m)v133 (total number of years of education).

. describe mv106 mv107 mv133 mv149

Variable name	Storage type	Display format	Value label	Variable label

mv106	byte	%8.0g	MV106	educational level
mv107	byte	%8.0g	MV107	highest year of education (at level in mv106)
mv133	byte	%8.0g	MV133	total number of years of education
mv149	byte	%8.0g	MV149	educational attainment

. tab mv107 mv149

highest year of education (at level in mv106)	educational attainment				higher	Total
-----+-----+-----						
no years completed at	239	0	0	0		239
1	366	2,470	0	2,071		4,907
2	1,172	4,915	0	2,237		8,324
3	1,636	10,124	0	9,066		20,826
4	2,455	11,970	0	1,004		15,429
5	5,842	13,488	0	2,797		22,127
6	0	0	4,211	371		4,582
7	0	12,840	0	185		13,025
8	0	0	0	111		111
-----+-----+-----						
Total	11,710	55,807	4,211	17,842		89,570

It seems to me that in the above table, the figure 4,211 should come under "incomplete secondary" and the figure 12,840 should come under "complete secondary". I see a similar result with mv133 and mv149 as well.

. tab mv133 mv149

total number of years of education	educational attainment				higher	Total
no educat	incomplet	incomplet	complete			

0	12,269	239	0	0	0	12,508
1	0	366	0	0	0	366
2	0	1,172	0	0	0	1,172
3	0	1,636	0	0	0	1,636
4	0	2,455	0	0	0	2,455
5	0	5,842	0	0	0	5,842
6	0	0	2,470	0	0	2,470
7	0	0	4,915	0	0	4,915
8	0	0	10,124	0	0	10,124
9	0	0	11,970	0	0	11,970
10	0	0	13,488	0	0	13,488
11	0	0	0	4,211	0	4,211
12	0	0	12,840	0	0	12,840
13	0	0	0	0	2,071	2,071
14	0	0	0	0	2,237	2,237
15	0	0	0	0	9,066	9,066
16	0	0	0	0	1,004	1,004
17	0	0	0	0	2,797	2,797
18	0	0	0	0	371	371
19	0	0	0	0	185	185
20	0	0	0	0	111	111
Total	12,269	11,710	55,807	4,211	17,842	101,839

I am seeing similar results for the individual recode file as well. Please let me know if I'm missing something.

Thank you for your help!

Subject: Re: Mismatch in educational attainment and education in single years
 Posted by [Bridgette-DHS](#) on Wed, 24 Jan 2024 18:31:37 GMT

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

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

You are asking about the Kenya 2022 survey. An erratum is being issued on the coding of education in this survey.

I suggest that you look at the relationship among v106, v107, v133, and v149 in the previous survey, the Kenya 2014 survey, because the relationship was ok in that survey. In the 2014 survey, you will see how v133 is supposed to be constructed as a recode of v106 and v107. In the IR file, do this:

. tab v107 v106, summarize(v133) means

Means of education in single years

highest year of highest educational level education primary secondary higher Total				
-----+-----+-----				
no years	0	8	12	8.7214076
1	1	9	13	9.2506473
2	2	10	14	10.554584
3	3	11	15	9.9238281
4	4	12	16	11.115502
5	5	.	17	5.3082022
6	6	14	18	6.2527945
7	7	.	19	7.0332318
8	8	.	20	8.0017804
-----+-----+-----				
Total	6.6739256	10.794648	14.333705	8.7562835"

Ignore the "total" row and column. You will see that each combination of v106 and v107 gives a value of v133. (There are two combinations that give v133=12 and two that give v133=14.) That is, v133 is a recode of v106 and v107.

If you enter "tab v106 v149" you will see that v149 is a recode of v106, which just distinguishes between whether the person completed the primary or secondary levels. That's based on the value of v107 within each level of v106.

That's all there is to it. v106 and v107 are the key variables; v133 and v149 are constructed variables and they were constructed incorrectly in the Kenya 2022 survey. You can either reconstruct them yourself, for your use, or wait for the next version of the Kenya 2022 data files.

Subject: Re: Mismatch in educational attainment and education in single years
Posted by [Navya](#) on Wed, 24 Jan 2024 18:50:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

Thank you for your quick response and all your help. I will construct the education attainment indicator based on the key variables (m)v106 and (m)v107.

Subject: Re: Mismatch in educational attainment and education in single years

Posted by fred.arnold@icf.com on Sat, 06 Apr 2024 19:06:28 GMT

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

I think that the problem may be that the NFHS education questions do not follow the DHS standard in that there are no questions about the level of education in NFHS surveys. The only education questions are "What is the highest grade you completed?" (Not the highest grade completed at that level, since there are no levels). I'm not sure how the NFHS-5 variables are created from the available education questions, but they were probably based on the highest grade completed, with decisions being made about what level that grade might reflect.
