
Subject: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)

Posted by [n.borgmann](#) on Wed, 14 Aug 2024 13:45:26 GMT

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

Dear DHS Team,

I am reaching out with a question regarding the DHS dataset for Kenya, specifically about the variables V107 (years of schooling) and V106 (education level) for the years 2022 and 2014.

In the 2022 dataset, there seem to be only a very small number of individuals with more than 8 years of schooling--just 7 individuals. Considering that Kenya's education system typically includes both primary and secondary schooling, which should exceed 8 years, I am puzzled by this finding.

I have seen discussions in the forum suggesting that there were issues with this variable in previous datasets and that the revised dataset will be released by the end of this week. However, the dataset for 2014 was referenced as being accurate. Upon reviewing the 2014 data, I found that the maximum value for years of schooling is also 8 years. This seems inconsistent with the data for variable V106, where many individuals fall into the third category, which is supposed to be higher than secondary education.

Could you please provide an explanation for these results? Is there a specific reason for this discrepancy, or could the data potentially be incorrect? Understanding this would be crucial for my work.

Thank you very much for your time and assistance. I greatly appreciate your help.

Best regards,

Nils

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)

Posted by [Bridgette-DHS](#) on Thu, 15 Aug 2024 14:07:51 GMT

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

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

You may be misinterpreting v107. That variable gives the number of years WITHIN the reported level, v106. The total number of years is v133, "education in single years". It is constructed by combining v106 and v107. In the Kenya 2014 survey, the following table has v106 for columns, v107 for rows, and v133 for cell values (ignore the Total row and column).

In the Kenya 2022 survey there was an error in the construction of the education variables. New versions of the files for that survey were put on the website just a few days ago and you should use them in place of the old files.

I can't account for illegal values, such as the 7 cases you found (if indeed they are illegal) in the KE 2022 survey. DHS files have remarkably few such values, but there are some. You may be able to find out whether and how they were altered by comparing with the new versions of the data files. (I can't do that because the internal data folder I use has not been updated yet).

Let us know if you have other questions.

File Attachments

1) [v107-v106.png](#), downloaded 321 times

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [n.borgmann](#) on Thu, 15 Aug 2024 14:12:31 GMT

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

Thank you very much for your detailed explanation. This clarifies everything, and I probably wouldn't have figured it out on my own. I greatly appreciate your help.

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [n.borgmann](#) on Fri, 16 Aug 2024 08:06:54 GMT

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

Thank you very much for the detailed response.

As I mentioned in another post, there was a mix-up on my end regarding the variable names.

Thank you so much for the help, my calculations make sense now!

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [pkaburi](#) on Fri, 03 Jan 2025 14:27:44 GMT

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

Dear DHS Users,

I have observed the trend of the Median years of education completed for both men and women over the last previous KDHS surveys. In the last two KDHS surveys, the median years of education completed among the high wealth quintile declined for both men (from 11.4 in 2014 to 10 in 2022) and women (from 11.2 in 2014 to 9.8 in 2022). This appears significant and I wonder what could be the explanation. Trends in the other wealth quintile show a positive trend. My analysis is based on statcompiler.

File Attachments

1) [Median years of education completed \[Men\].pdf](#), downloaded

91 times

2) [Median years of education completed \[Women\].pdf](#), downloaded

87 times

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [Bridgette-DHS](#) on Mon, 06 Jan 2025 14:45:59 GMT

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

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

This is a very interesting observation, but before trying to understand it, I first need to know how you got the estimates of the medians.

Did you use hv108 in the PR file? If so, did you have lower or upper cutoff ages? Tables 2.13.1-2 in the final report on the 2014 survey, for example, include ages 6+, with no upper age cutoff. Or did you use v133 in the IR file, which would apply to age 15-49 (for women) and mv133 in the MR file, which would apply to age 15-54 (for men)? Did you modify the upper age limit for men to match that of women?

I can suggest a strategy to understand this pattern, which I agree is unexpected, if you will clarify how you got the numbers.

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [pkaburi](#) on Mon, 06 Jan 2025 15:13:59 GMT

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

Many thanks for your response. As reported in the database, I got the figures directly from the DHS stat compiler. I did not do any extra analysis but did a trend analysis over time for all previous KDHS.

Regards

Subject: Re: Issues with Education Data in DHS Dataset for Kenya (2022 and 2014)
Posted by [Bridgette-DHS](#) on Fri, 24 Jan 2025 14:27:29 GMT

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

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

I apologize for the long delay in this response. I have been able to confirm that these medians were calculated from hv108 (in the PR file) in both surveys. hv108 is constructed hv106 (highest level attended) and hv107 (years attended at that level). In the following table, using the 2014

survey, hv106 is the column variable, hv107 is the row variable, and hv108 is the constructed variable, in each cell for the total number of years of schooling (ignore the totals row and the totals column).

```
. tab hv107 hv106, summarize(hv108) means
```

Means of education completed in single years

highest	highest educational level			
year of	attained			
education	primary	secondary	higher	Total
completed				
0	0	8	12	3.5867717
1	1	9	13	4.831788
2	2	10	14	6.9003889
3	3	11	15	6.5741577
4	4	12	16	9.3242178
5	5	13	17	5.2350961
6	6	14	18	6.3245768
7	7	.	19	7.0337612
8	8	.	20	8.0081755
9	.	.	21	21
10	.	.	22	22
don't know	98	98	98	98
-----+-----+-----				
Total	4.8757119	10.761662	14.517886	6.9356705

Here is the corresponding table for the 2022 survey:

```
. tab hv107 hv106, summarize(hv108) means
```

Means of education completed in single years

highest	highest educational level			
year of	attained			
education	primary	secondary	higher	Total
completed				
0	0	8	12	5.6417091
1	1	9	13	5.6580055
2	2	10	14	7.7659006
3	3	11	15	8.8481961
4	4	12	16	10.427846
5	5	13	17	5.2305875

6	6	14	18	6.6000514
7	7	.	19	7.0442526
8	8	.	20	8.0260387
9	.	.	21	21
10	.	.	22	22
11	.	.	23	23
12	.	.	24	24
don't know	98	98	98	98
-----+-----+-----				
Total	5.3123704	11.025853	15.361057	8.1386855

There were some changes between 2014 and 2022, which added 23 and 24 years to the distribution, but only a total of 9 people in the 2022 survey have those values so that is not an issue. The "don't know" cases were omitted from the calculations.

I believe that what you see for median years of schooling in the top quintile could be understood by disaggregating birth cohorts of women and men, or equivalently age groups. It's an interesting question, as I said, but unfortunately I have not been able to find time to do the kind of disaggregation that I just described. Perhaps other users can make suggestions or follow up.