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 |
   year of |   highest educational level
 education |          attained
 completed |   primary  secondary     higher |     Total
-----------+---------------------------------+----------
         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 |
   year of |   highest educational level
```

```
education |          attained
completed |   primary  secondary     higher |     Total
-----------+--------------------------------+----------
         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.