Subject: Re: Kenya DHS weights

Posted by Bridgette-DHS on Tue, 31 Oct 2023 13:25:18 GMT

View Forum Message <> Reply to Message

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

First, weighted and unweighted frequencies are always different, sometimes by large amounts. If, in the PR file, you enter the following lines, you will see that the weights for children (age<18) are usually much less than 1 (ignoring the factor of one million).

gen wt=hv005/1000000 collapse (mean) wt, by(hv105)t graph hbar wt, over(hv105) yline(1)

In this survey, apparently, fertility tended to be higher (that is, there were more children) in the strata (geographic areas) that were over-sampled. In order for the estimates to be unbiased, the weights for these strata tended to be less than one (ignoring the factor of one million).

Second, regarding the education variables, I suggest is that you read the questions, the codes, and the tables and text in the report very carefully. I agree that what you are seeing is hard to believe. I looked at the data too. There are many inconsistencies. For examples, among the 3-year old (hv105=3) I see a child who, according to hv126, attended secondary school the previous school year; there are 5 3-year olds who are in primary school this year AND were in primary school the previous year. Perhaps primary school includes day care?

The school attendance variables (hv121-hv129) were probably not edited as much as they could have been. These variables tend to be somewhat survey-specific, because of differences in standards (e.g. the starting age) and definitions from one country to another and even over time in the same country. You may want to edit the responses yourself for consistency.

Usually it would help to do "tab hv124 hv122" to see how single years line up with levels. However, when I do that, I see a possible coding error involving hv124 for single years 7-12. We will look into this further and may add another post on that.