

---

Subject: Re: Weighting variables in DHS India data (1992 and 1998)

Posted by [Bridgette-DHS](#) on Tue, 01 Sep 2015 16:41:48 GMT

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

---

Here is another response from Tom Pullum:

Sampling weights are inflation or deflation factors that have an average value of 1. DHS moves the decimal point six places to the right, that is, multiplies the weight by 1,000,000 (one million). Thus "1148330" is actually "1.148330". The reason for doing this is the same as the reason why percents have a factor of 100, fertility and mortality rates often have a factor of 1000, and maternal mortality rates have a factor of 100,000. It's just to have many significant digits without having to worry (or worry very much) about a decimal point. A weight of 1.148330 for each child in the cluster just means that that the probability that a child in that cluster would appear in the sample was a little less than the average for the whole country, and therefore, to compensate, those children get a weight greater than 1. It doesn't really have anything to do with the population of the district. It is impossible, just from the weights, to estimate the population at any level of aggregation.

When you collapse, you are usually calculating a mean. For example, if you wanted to calculate the mean of some variable x in the district, you would use "collapse (mean) x [iweight=v005/1000000]" . This would give you an estimate of the mean of x in the district, corrected or adjusted for the weights. If you did not use the weights, the mean of x would be biased toward the oversampled children. (With collapse, the default is the mean, so you would get the same thing with just "collapse x [iweight=v005/1000000]".)

I don't know what you are doing with "collapse(rawsum)v005/1000000 , by (district)". You have omitted something, and I don't just mean spaces.

"(rawsum)" indicates a sum, not a mean, and it ignores the weights. I think "sum" is the only statistic that can be prefaced by "raw". There is no "rawmean", for example.

I think you may be making this more complicated than necessary. I would say that you need to use the weights for any collapses and for any estimation commands. Avoid having the numbers come out 1,000,000 times larger than they should be. That's all.

Let me know if you still have questions.