Subject: Education - Net Attendance Ratio (Primary)
Posted by djohnson on Wed, 02 Aug 2017 21:54:16 GMT

I am having trouble replicating the Net Attendance Ratio for Primary Education that was presented in the DHS Tanzania 2015/16 Final Report. I was originally trying to just calculate confidence intervals, but now I am worried I have missed something along the way since the rate itself doesn't match (I get 73.979%, DHS report shows 75.7%).

I am using the TZPR7HFL.DTA dataset for Stata, variable hv121 to represent attending, hv122 for the education level, and hv105 for age.

I am also using:
weight --> HV005/1,000,000
cluster --> HV021 Primary sampling unit
strata --> HV023 Stratification used in sample design

More or less (with minor background formatting/variable generation), I'm doing the following:

```
      gen net attend = 0
 replace netattend = 100 if attending==1 & inrange(age,7,13) & level==1

 svyset cluster [pweight=hhweight], strata(strata)
 svy, subpop(if inrange(age,7,13)) : mean netattend
 estat size, obs
```

Thanks!

Subject: Re: Education - Net Attendance Ratio (Primary)
Posted by Trevor-DHS on Wed, 16 Aug 2017 01:26:51 GMT

The calculation of the net attendance ratios is actually a bit more complicated than it looks. When calculating the attendance ratios, DHS actually calculates the age of each child at the time of the start of the academic year. To do this, DHS merges the century month code of the date of birth for any child listed in the women's birth history to the entry in the PR datasets, matching B16 in the BR dataset with HVIDX in the PR dataset. DHS then calculates the age at the start of the school year (I believe the school year starts in January each year in Tanzania) by subtracting the century month code of the date of birth from the century month code of the start of the school year (CMC = 1381 = (2015-1900)*12+1, or 1383 = (2016-1900)*12+1 if the interview was in January 2016 or later). For any child that did not have a century month code of the date of birth in the women's data, a randomly imputed month is chosen, and this is used to calculate the age at the start of the school year. This age is then used, not the current age of the child at the time of the survey.