
Subject: KDHS 2022: Table 2.12 School attendance ratios

Posted by [geok](#) on Wed, 21 Aug 2024 21:27:44 GMT

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

Dear DHS staff,

I'm trying to match Table 2.12 of school attendance ratios, using GitHub code shown below. I used the revised KDHS 2022 dataset that was released on August 13th.

```
use "KEBR8CFL.DTA", clear
```

```
* keep only the variables we need
```

```
keep v001 v002 v003 b3 b16
```

```
* drop if the child in the birth history was not in the household or not alive
```

```
drop if b16==0 | b16==.
```

```
* rename key variables for matching
```

```
rename b16 hvidx
```

```
rename v001 hv001
```

```
rename v002 hv002
```

```
* sort on key variables
```

```
sort hv001 hv002 hvidx
```

```
* if there are some duplicates of line number in household questionnaire, we need to drop the duplicates
```

```
gen dup = (hv001 == hv001[_n-1] & hv002 == hv002[_n-1] & hvidx == hvidx[_n-1])
```

```
drop if dup==1
```

```
drop dup
```

```
* re-sort to make sure still sorted
```

```
sort hv001 hv002 hvidx
```

```
* save a temporary file for merging
```

```
tempfile tempBR
```

```
save `tempBR'
```

```
* use the PR file for household members for the NAR and GAR indicators
```

```
use "KEPR8CFL.DTA", clear
```

```
* merge in the date of birth from the women's birth history for the household member
```

```
merge 1:1 hv001 hv002 hvidx using `tempBR'
```

```
* there are a few mismatches of line numbers (typically a small number of cases) coming from the BR file, so let's drop those
```

```
drop if _merge==2
```

```
* restrict to de facto household members age 5-24, and drop all others
```

```
keep if hv103==1 & inrange(hv105,5,24)
```

```
* now we calculate the child's age at the start of the school year
```

* but first we have to specify the month and year of the start of the school year referred to in the survey

* example, for Zimbabwe 2015 survey this was January 2015

```
global school_start_yr = 2022
```

```
global school_start_mo = 1
```

* also need the age ranges for primary and secondary

```
global age_prim_min = 6
```

```
global age_prim_max = 13
```

```
global age_sec_min = 14
```

```
global age_sec_max = 17
```

* produce century month code of start of school year for each state and phase

```
gen cmcSch = ($school_start_yr - 1900)*12 + $school_start_mo
```

```
replace cmcSch = cmcSch+12 if hv008 >= cmcSch+12
```

* calculate the age at the start of the school year, using the date of birth from the birth history if we have it

```
gen school_age = int((cmcSch - b3) / 12) if b3 != .
```

* Impute an age at the beginning of the school year when CMC of birth is unknown

* the random imputation below means that we won't get a perfect match with the report, but it will be close

```
gen xtemp = hv008 - (hv105 * 12) if b3 == .
```

```
gen cmctemp = xtemp - int(uniform()*12) if b3 == .
```

```
replace school_age = int((cmcSch - cmctemp) / 12) if b3 == .
```

* Generate variables for whether the child is in the age group for primary or secondary school

```
gen prim_age = inrange(school_age,$age_prim_min,$age_prim_max)
```

```
gen sec_age = inrange(school_age,$age_sec_min,$age_sec_max)
```

* create the school attendance variables, not restricted by age

```
gen prim = (hv122 == 1)
```

```
gen sec = (hv122 == 2)
```

* set sample weight

```
cap gen wt = hv005/1000000
```

* For NAR we can use this as just regular variables and can tabulate as follows, but can't do this for GAR as the numerator is not a subset of the denominator

* NAR is just the proportion attending primary/secondary school of children in the correct age range, for de facto children

```
gen nar_prim = prim if prim_age == 1
```

```
gen nar_sec = sec if sec_age == 1
```

```
lab var nar_prim "Primary school net attendance ratio (NAR)"
```

```
lab var nar_sec "Secondary school net attendance ratio (NAR)"
```

* tabulate primary school attendance

```
tab hv104 nar_prim [iw=wt] , row
```

```
tab hv025 nar_prim [iw=wt] , row
```

```
tab hv270 nar_prim [iw=wt] , row
* tabulate secondary school attendance
tab hv104 nar_sec [iw=wt] , row
tab hv025 nar_sec [iw=wt] , row
tab hv270 nar_sec [iw=wt] , row
```

Can you please help?

Subject: Re: KDHS 2022: Table 2.12 School attendance ratios
Posted by [Bridgette-DHS](#) on Thu, 22 Aug 2024 16:42:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

The problem is with your specification of school_start_yr and school_start_mo. You used 2022 and 1, respectively. That was a good guess, but the tables are based on 2021 and 7, a modification because of Covid.

I looked at the CSPpro code for chapter 2 and it has the following notes:

```
{cm: KNBS explained survey academic year in 2022 started in July 2021 and ended in April 2022.
The survey started in January 2022.}
{ The 2022 academic year started on May 2022 and will finish on December 2022. This start
months for academic year were affected by }
{ the pandemic. Schools will open normal schedule (as before the pandemic) in 2023, from
January to December }
yeareduc = 2021; { done - Survey academic year. Use first year if survey goes across two
years }
mntheduc = 07; { done - Month when survey academic year starts}
```

For other users who try to reproduce the school attendance ratios--the most common reason for not getting a match is that you do not have the "correct" values of these two numbers.
