Subject: School attendance Mozambique 2022
Posted by Simon on Fri, 24 Jan 2025 15:39:16 GMT
View Forum Message <> Reply to Message

Hi there,
I'm aiming at the NARs, GARs and GPIs for school attendance in Mozambique based on the 2022 survey - basically table 2.12.

I changed the year and start month in the code download from GitHub to 2022 and 1, and that doesn't seem to match. I actually tried all the combinations of years and months from 2021 1 to 2022 12 - and none of them seem to match!

I think I'm missing something else, and was hoping you might point me in the right direction.

Thanks a lot! Simon

P.S. Below is the GitHub code:

```
 /***************************************************** ****************************************
Program:    PH_SCHOL.do
Purpose:    Code to compute school attendance indicators
Data inputs:   BR and PR dataset
Data outputs:  coded variables
Author:    Trevor Croft, modified by Shireen Assaf for this project
Date last modified: May 6, 2020 by Shireen Assaf
                    December 1, 2021 by Trevor Croft

Note:    To produce the net attendance ratios you need to provide country specific information on the year and month of the school calendar and the age range for school attendance. See lines 63-73. You can obtain this information for each country from the UNESCO webiste: http://data.uis.unesco.org/. This would be under "Education" and then "Other policy relevant indicators". Scroll to the bottom of the list to obtain the school ages from "Offical entrance age to each ISCED level of education" and the school calendar from "Start and end of the academic year".
 ***************************************************** ****************************************/

 /*---------------------------------------------------------- ------------------
Variables created in this file:

ph_sch_nar_prim   "Primary school net attendance ratio (NAR)"
ph_sch_nar_sec   "Secondary school net attendance ratio (NAR)"
ph_sch_gar_prim   "Primary school gross attendance ratio (GAR)"
ph_sch_gar_sec   "Secondary school gross attendance ratio (GAR)"
ph_sch_nar_prim_*_gpi "Gender parity index for NAR primary"
ph_sch_nar_sec_*_gpi "Gender parity index for NAR secondary"
ph_sch_gar_prim_*_gpi "Gender parity index for GAR primary"
```

```
ph_sch_gar_sec_*_gpi "Gender parity index for GAR secondary"
 ------------------------------------------------------------ ----------------*/

* For net attendance rates (NAR) and gross attendance rates (GAR) we need to know the age of children at the start of the school year.
* For this we need to get date of birth from birth history and attach to children's records in the PR file.
* open the birth history data to extract date of birth variables needed.
use "C:\Users\Simon Cresswell\Desktop\Cornerstone\Mozambique\Dropbox files\DHS\DHS 2022\DHS Database\4. Birth History\MZBR81FL.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
save tempBR, replace

* use the PR file for household members for the NAR and GAR indicators
use "C:\Users\Simon Cresswell\Desktop\Cornerstone\Mozambique\Dropbox files\DHS\DHS 2022\DHS Database\7. Household Members\MZPR81FL.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 rom 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
* example, for Zimbabwe 2015, the age range is 6-12 for primary school and 13-18 for secondary school
global age_prim_min = 6
global age_prim_max = 12
global age_sec_min = 13
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 seconary 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


* Program for calculating NAR or GAR
* NAR just uses a mean of one variable
* GAR uses a ratio of two variables

* Program to produce NAR or GAR for background characteristics (including total) for both sex, combined and separately
cap program drop nar_gar
program define nar_gar
  * parameters
  * type of rate - nar or gar
  * type of schooling - prim or sec
  * background variable for disaggregation

  * generates variables of the following format
  * ph_sch_`rate'_`sch'_`backvar'_`sex'
  * e.g. ph_sch_nar_prim_total_0
  * or   ph_sch_gar_sec_hv025_2
  * sex: 0 = both sexes combined, 1=male, 2=female

  * type of rate - nar or gar
  local rate `1'
  if "`rate'" != "nar" & "`rate'" != "gar" {
 di as error "specify type of rate as nar or gar"
 exit 198
  }
  * type of schooling - prim or sec only
  local sch `2'
  if "`sch'" != "prim" & "`sch'" != "sec" {
 di as error "specify schooling as prim or sec"
 exit 198
  }
  * name of background variable
  local backvar `3'
  * do for total = 0, and each sex male = 1, female = 2
  foreach sex in 0 1 2 {
    if `sex' == 0 local select 0==0 /* always true */
    else          local select hv104==`sex'
 if "`rate'" == "nar" { /* Net Attendance Rate (NAR) */
   mean `sch' [iw=wt] if `select' & `sch'_age == 1, over(`backvar')
   * results matrix for mean - used for NAR
   mat x = e(b)
```

```
 }
 else { /* Gross Attendance Rate (GAR) */
      ratio `sch' / `sch'_age [iw=wt] if `select', over(`backvar')
   * results matrix for ratio - used for GAR
      mat x = r(table)
 }
 * generate the output variable we will fill
    gen ph_sch_`rate'_`sch'_`backvar'_`sex' = .
 * get all of the characteristics of the background variable
    cap levelsof `backvar'
    local ix = 1
    local lev `r(levels)'
 * loop through the characteristics and get the result from matrix x
    foreach i in `lev' {
   * capture the result for this characteristic
      replace ph_sch_`rate'_`sch'_`backvar'_`sex' = 100*x[1,`ix'] if `backvar' == `i'
      local ix = `ix' + 1
    }
 * label the resulting variable
 local schooling primary
 if "`sch'" == "sec" local schooling secondary
 local sexlabel both sexes
 if `sex' == 1 local sexlabel males
 if `sex' == 2 local sexlabel females
 lab var ph_sch_`rate'_`sch'_`backvar'_`sex' "`rate' for `schooling' education for background
characteristic `backvar' for `sexlabel'"

    * Tabulating indicators by background variables and exporting estimates to excel table
Tables_schol.xls
 * the tabulations will provide the estimates for the indicators for the total, males, and females for
the background variable
    tabout `backvar' using Tables_schol.xls, sum cells(mean ph_sch_`rate'_`sch'_`backvar'_`sex')
ptotal(none) append
  }

  * gender parity index for a rate for a characteristic - female (2) rate divided by male (1) rate
  gen ph_sch_`rate'_`sch'_`backvar'_gpi = (ph_sch_`rate'_`sch'_`backvar'_2 /
ph_sch_`rate'_`sch'_`backvar'_1)
  lab var ph_sch_`rate'_`sch'_`backvar'_gpi "gender parity index for `rate' for `schooling' education
for background characteristic `backvar'"

  * Tabulating the GPI indicator by background variable and exporting estimates to excel table
Tables_schol.xls
  * the tabulations will provide the estimates for the GPI indicator for the background variable
  mean ph_sch_`rate'_`sch'_`backvar'_gpi, over(`backvar')
  tabout `backvar' using Tables_schol.xls, sum cells(mean ph_sch_`rate'_`sch'_`backvar'_gpi) f(2)
ptotal(none) append
```

```
end

* create total background characteristic
gen total = 0
lab var total "total"
lab def total 0 "total"
lab val total total

cap erase Tables_schol.xls

* Caculate indicators and save them in the dataset
nar_gar nar prim total /* NAR primary   - total population */
nar_gar nar prim hv025 /* NAR primary   - urban/rural */
nar_gar nar prim hv024 /* NAR primary   - region */
nar_gar nar prim hv270 /* NAR primary   - wealth index */

nar_gar nar sec  total /* NAR secondary - total population */
nar_gar nar sec  hv025 /* NAR secondary - urban/rural */
nar_gar nar sec  hv024 /* NAR secondary - region */
nar_gar nar sec  hv270 /* NAR secondary - wealth index */

nar_gar gar prim total /* GAR primary   - total population */
nar_gar gar prim hv025 /* GAR primary   - urban/rural */
nar_gar gar prim hv024 /* GAR primary   - region */
nar_gar gar prim hv270 /* GAR primary   - wealth index */

nar_gar gar sec  total /* GAR secondary - total population */
nar_gar gar sec  hv025 /* GAR secondary - urban/rural */
nar_gar gar sec  hv024 /* GAR secondary - region */
nar_gar gar sec  hv270 /* GAR secondary - wealth index */

erase tempBR.dta
```

Subject: Re: School attendance Mozambique 2022
Posted by Janet-DHS on Mon, 27 Jan 2025 17:46:51 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:
The month and year you need for these rates can vary from one survey to another. I checked the CSPro code for the construction in this survey and you should be ok with year=2022 and month=2 (see the lines below).

```
  yeareduc = 2022;   { Survey academic year. Use first year if survey goes across two years }
  mntheduc = 02;     {  Month when survey academic starts. Ask for it to SM  }
  cmceduci = cmcode( mntheduc, yeareduc );
```

However, you say you tried a range that would have included 2/2022. For some children the CMC of birth is not known and is randomly imputed within a 12-month range. If you still do not match the table, it will be because of that random component. It will be impossible to match the random

component exactly, but you should get very close with a similar imputation procedure. This is not an issue in all surveys. Sorry about that....
seed(101); { to initiate randomization for members age at the beginning of school year }

{ impute an age at the beginning of the school year when CMC of birth unknown }
xtemp = HV008 - HV105*12;
cmctemp = random(xtemp-11, xtemp);
ageatsch = int((cmceducf-cmctemp) / 12);

Subject: Re: School attendance Mozambique 2022
Posted by Simon on Mon, 27 Jan 2025 22:48:12 GMT
View Forum Message <> Reply to Message

Superb! Thanks very much for the quick reply!
Best
Simon
