Subject: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 31 Oct 2023 10:53:41 GMT
View Forum Message <> Reply to Message

Dear all,

I am looking at school enrollment, progression and repetition for children in Kenya - I am looking at these indicators grade-wise. I have a few questions:

1. For school enrollment - I am looking at hv121 and hv125 (member attended school current and previous year respectively) - now for Grade 1 I see there are respondents who said "yes" to hv125 (attended school last year) - what does it mean? Does it mean these children were in pre-primary or early childhood education? I am asking because the number is high.

2. Secondly, to make calculations easier - I have re-coded secondary-level grades and higher level grades. Secondary-level grades have been coded as 9,10,11,12,13,14 and similarly higher-education grades have been coded from 13-19. Nonetheless, I am using the following code:

tab agrdprvyr_i scheprvyr_i [iweight = hv005/1000000]

agrdprvyr_i is previous year grades (coded as hv127 in DHS) and scheprvyr_i is school enrollment last year (coded as hv125 in DHS)

My question is that when I use weights, why is the "number of observations" decreasing? Secondly the data is showing that only 1621 individuals were enrolled in grade 12 last year - but this is a very small number, and doesn't seem right. Am I doing anything wrong with the weights? Please help.

I have attached snapshots of both unweighted and weighted data

Best,
Ritapriya

File Attachments
1) school-enrollment unweighted.PNG, downloaded 100 times
2) school-enrollment weighted.PNG, downloaded 102 times

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.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 31 Oct 2023 13:50:09 GMT
View Forum Message <> Reply to Message

Hi

Thank you for the response

I understood your point, but with regards to weights, what should I do? Does it mean that what I did was correct, and that there are 1621 individuals in Grade 12 previous year, or should I use a different formula. Could you please clarify - I do understand that the weights of children is less than 1 - but wanted to ask what should I do about it - and how should I apply weights then

My current formula is tab hv121 hv127 [iweight = hv005/1000000]

Best
Ritapriya

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 31 Oct 2023 13:58:26 GMT
View Forum Message <> Reply to Message

Sorry

So does this mean I should just use tab hv121 hv127 [iweight = hv005] - and not divide by 1 million? - In that case, the frequencies increase a lot - attached snapshot

Best
Ritapriya

File Attachments
1) school enrollment.PNG, downloaded 92 times

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Tue, 31 Oct 2023 14:56:40 GMT
View Forum Message <> Reply to Message

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

The weight includes a factor of 1 million just to get rid of the decimal point. The mean of hv005 (ignoring that factor) in the HR file, with households as units, is 1.

You should always use weights to get unbiased estimates of means, proportions, percentages, ratios, rates, etc. You are using the correct formula.

The frequencies in the sample, unweighted or weighted, are not important--as I said, it's the means, etc., that are important, and you are ok with them.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 31 Oct 2023 15:13:49 GMT
View Forum Message <> Reply to Message

Okay thank you - I am going forward with: tab hv125 hv127 [iweight = hv005/1000000] then? - to tabulate previous year grade and enrolment status

Could you please also clarify what you meant by primary school involving day care? I have checked the report and coding - before primary level - there's only one level: no education/preschool/early childhood (this is coded as 0, primary level as 1, secondary as 2, higher as 3)

So I am assuming if grade 1 goers have "1" or "yes" to hv125 (member attended school last year) - then they were in early childhood/preschool education?

Also if children are in primary school this year and previous year, in the same grade - I believe that shows grade repetition as well

Best,
Ritapriya

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Fri, 03 Nov 2023 12:23:56 GMT
View Forum Message <> Reply to Message

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

I hope you recognize that grades are numbered within levels. In KEPR8B you can get the combinations of grade and level in the PREVIOUS YEAR with "tab hv127 hv126". These combinations are converted to education in single years (previous year), hv128. To see this, enter "tab hv127 hv126, summarize(hv128) means." This will give the value of hv128 within each combination of hv127 and hv126.

For the CURRENT YEAR you can do "tab hv123 hv122" and "tab hv123 hv122, summarize(hv124) means".

These tabulations do not use weights because I just want to describe the coding pattern. For actual analysis you would use weights.

You will see that year 7 can be either the 7th year of primary or the 1st year of secondary, and year 8 can be either the 8th grade of primary or the 2nd year of secondary. I am suspicious that there was a coding error somewhere in the hv12* variables, affecting years "single years" 7-12, and have asked the data processing staff to look into this. We will post more after hearing from them. Until then, I recommend caution.

To be clear, I agree that primary school is not "supposed to" include daycare. When I mentioned that this could happen I was describing potential errors or misinterpretations by the respondents or interviewers. I wanted to understand how very young children could be reported as attending primary school.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Mon, 06 Nov 2023 10:21:58 GMT
View Forum Message <> Reply to Message

Hi,

So yes, I got your point. I am using KEPR8BDT. However, I see that for last year grades, there is no one in grade 7 or 8 who is classified to be in secondary education - I have attached a snapshot of this. I have re-coded secondary grades to 9,10,11 and 12. Kenya follows a 8-4-4 system - so 8 years primary and then 4 years secondary - hence this aligns with Kenyan education system. However, it is also true that when I use education in single years - as in hv128 - I see that grade 7/8 can belong to either primary or secondary - so what does it mean? Does it mean that even though grade 7/8 are coded as primary grades, there's a possibility that they may be under secondary?

Best
Ritapriya

File Attachments
1) School-enrollment previous year.PNG, downloaded 89 times

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Mon, 06 Nov 2023 14:32:01 GMT
View Forum Message <> Reply to Message

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

The data processing staff are revising the recodes for these variables. I hope to post them soon.

Subject: Re: Kenya DHS weights
Posted by sokiya on Mon, 06 Nov 2023 15:57:42 GMT
View Forum Message <> Reply to Message

Let me jump into this discussion since I am Kenyan and privy to the recent changes in the education system.
Actually, those are not coding errors. The education is transitioning fully (in my opinion from next year) from 8-4-4 system to 2-6-3-3-3. So it possible that during the data collection period, one would say that s/he in grade 6 and 7 and in junior secondary and another in the same grade but in primary. The last lost of students in grade 8 in primary have just sat their KCPE

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Mon, 06 Nov 2023 18:15:35 GMT
View Forum Message <> Reply to Message

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

Several variables giving single years of education need to be recoded. They were previously coded with 6 years of primary and 6 of secondary. However, as you say, Kenya has 8 and 4, respectively. For example, hv128 in the PR file was constructed as

```
gen     hv128=   hv127 if hv126==1
replace hv128= 6+hv127 if hv126==2
replace hv128=12+hv127 if hv126==3
```

However, it should have been calculated as follows, with 8 rather than 6 in the second line:

```
gen     hv128=   hv127 if hv126==1
replace hv128= 8+hv127 if hv126==2
replace hv128=12+hv127 if hv126==3
```

The variables affected--that is, that need to be recoded--are hv108, hv124, hv128, v133, v715, and mv715. The next update of the files will include corrected versions of these variables, but you can fix them now.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 07 Nov 2023 09:43:32 GMT
View Forum Message <> Reply to Message

Dear all,

Thank you so much for all the suggestions and corrections - I would definitely keep them in mind for analyzing education data from the Kenya DHS 2022. Thank you also for letting me know that the Kenyan education system is changing.

I had a few questions:

Just to confirm, primary school is 8 years of education, secondary schooling is 4, then how many years are higher education? Also by this format Grade 7/8 are in primary level then? - do you know when the updated data will be released?

Best,
Ritapriya

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Tue, 07 Nov 2023 14:44:15 GMT
View Forum Message <> Reply to Message

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

Yes, in an 8+4 breakdown, grades 7 and 8 would be grouped with primary. I would expect that the changes in the primary and secondary breakdown would not affect the standard number of years for post-secondary.

DHS staff cannot respond to questions about country-specific education systems. Perhaps other users can help. You may also be able to find a Kenyan or UN website with such information.

I don't know when the updated files will be released, but when it happens, a data alert will be sent out to users who downloaded the previous version. Only the recodes of single years of schooling are affected and you can revise them yourself.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Fri, 10 Nov 2023 10:13:05 GMT
View Forum Message <> Reply to Message

Hello, thanks a lot.

I am just wondering that beyond 9,10,11,12 - there are two other grades (13/14) - which I believe are coded as secondary - so what to do with that?

I wanted to confirm two more things:

1. hv121 and hv125 represent school enrollment in current and previous school year right?

2. Secondly, I wanted to confirm - I saw a variable coded as hv129 - it gives the number of students repeated, dropouts and progressed/advanced. Could you please tell me the STATA codes for calculating these? The reason I am asking this is because my definitions of repeats, progression and dropouts are the same as DHS (as I verified from DHS Recode Manual). Now for dropouts my numbers are same as DHS - but it doesn't match for repeats - hence I wanted to know the code for that.

I followed your advice, and have changed education in single years - 8 years primary and 4 years secondary. I renamed education single year previous year to "edusyr_prvyr" and education single year current year to "edusyr_cr". Now I did tab edusyr_prv hv129 and I did tab edusyr_cr hv129 - as you can see the repeat numbers are different in both these commands - but why so? This shouldn't happen - repeat numbers should stay same I believe if I use either current or previous year.

Secondly, if you see grade 7 and 8 particularly - there are very large numbers of students who are repeating - while this is believable, it doesn't match with mine, wondering why.

I have attached snapshots for:

1. tab edusyr_prvyr hv129 [education single year previous year - edusyr_prvyr]
2. tab edusyr_cr hv129 [education single year current year - edusyr_cr]
3. tab edusyr_cr repeat [education single year current year - edusyr_cr] - this is my calculation of repeats

File Attachments
1) hv129_previous year DHS.PNG, downloaded 90 times
2) hv129_current year DHS.PNG, downloaded 84 times
3) hv129_author calculation.PNG, downloaded 86 times

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Fri, 10 Nov 2023 13:43:28 GMT
View Forum Message <> Reply to Message

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

These are good questions, but we cannot answer them at this time. The construction of the education variables in this survey is currently being reviewed by the DHS data processing staff. We will respond as soon as possible.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Thu, 16 Nov 2023 13:29:56 GMT
View Forum Message <> Reply to Message

Hi,

Thank you,

Just one thing - i am trying to look at dropouts right. Now i have calculated dropouts as those who said "yes" to attending school past year but not this year - however there are people who might have graduated right after class 12 (8+4 school system) - so I was wondering how to figure out how many graduated - should I do tab hv128 (past year grade - single education years) sh19aa (reason for not attending school) or should I do hv128 hv106 (hv106 - highest grade attended)

Best
Ritapriya

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Mon, 20 Nov 2023 13:07:14 GMT

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

If someone attended grade 12 last year (hv126=2 and hv127=4), and is reported with "secondary" as the highest completed level (hv106=2), then they must have graduated last year. I think you can identify last year's high school graduates with this combination of hv126, hv127, and hv106. Other combinations involving current attendance (hv121-hv123) should identify dropping out or repeating grade 12 or continuing to post-secondary schooling.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 21 Nov 2023 07:34:04 GMT

Clarify one thing

There are two variables that can give me school attendance in current/previous year - one is hv121/hv125 (school attendance status) and I believe the other is hv124/hv128 (education single years for each each year). As per our last discussion, you asked me to construct the hv128 variable as follows:

gen hv128= hv127 if hv126==1
replace hv128= 8+hv127 if hv126==2
replace hv128=12+hv127 if hv126==3

This was done to ensure 8 years of primary and 4 years of secondary school. I have done the same - I named it hv1281 so that original variable stays intact. Now, if I do tab hv1281 or tab hv1281 scheprvyr_i (latter one is the hv125 variable) - the numbers are same. As you can see, there are 5,113 students in grade 1 - so this is fine - I have attached screenshots for both commands.

However, as you can see, when I do hv1281 hv106 - there are students of grade 1 who have replied either no education or completed secondary education to hv106 - So the 5,113 individuals in Grade 1 comprise these individuals too? This is what I am confused about - 5,113 individuals said yes to school enrollment in previous year in grade 1 - but they also include those who said no education and completed secondary education to hv106? - so how many people are enrolled in grade 1 - 5,113 or 5,064?

Secondly, if that's the case then the 1,272 individuals who were in grade 12, secondary and said they completed education last year - can I really count them as graduated? Especially since there were also students in last year grade 12 who said they completed higher education. Plus, if I do hv1281 hv129 - the numbers for advanced, dropout, and repeats don't match with the 1,272 (number of students who are supposedly graduates as per hv1281 hv106) I have attached

screenshots of hv1281 hv106 and hv1281 hv129.

To avoid confusion - I was looking at hv1281 sh19aa for graduates. From the 939 people in class 12 last year who dropped out (that is, not attending this year - you can see from hv1281 hv129), I am subtracting those in class 12 last year who said they stopped schooling as they have completed school (i get this number from hv1281 sh19aa).

I understand education variables are going through re-checking - but please do help at the moment - I may be incorrect in my approach

Best,
Ritapriya

File Attachments

1) tab hv1281.PNG, downloaded 85 times
2) tab hv1281 scheprvyr_i.PNG, downloaded 85 times
3) tab hv1281 hv106.PNG, downloaded 88 times
4) tab hv1281 hv129.PNG, downloaded 87 times

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Tue, 21 Nov 2023 14:19:35 GMT
View Forum Message <> Reply to Message

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

I cannot give further recommendations for the recoding of the education variables in the Kenya 2022 survey. I do not want to risk saying something that will differ from the review. When the review has been completed, we will post the recodes and they will be included in the next update of the data files.

Subject: Re: Kenya DHS weights
Posted by Ritapriya Bandyopadhyay on Tue, 28 Nov 2023 10:11:40 GMT
View Forum Message <> Reply to Message

Thank you so much, and yes, I agree

Also, could you please let me know if there's any variable in the DHS HH module where I can get the total number of births in a year? I am actually trying to find out the proportion of births in the lowest quintile - q1 births/total births in previous year.

If not, then should I go to women's module?

Best
Ritapriya

Subject: Re: Kenya DHS weights
Posted by Bridgette-DHS on Tue, 28 Nov 2023 16:49:06 GMT
View Forum Message <> Reply to Message

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

If you open the BR file and enter "tab b2 v190 [iweight=v005/1000000], row" you will get what I believe you are looking for, going back in single calendar years. In the BR and KR files, b2 is the calendar year of birth. Note that the household's wealth quintile at the time of the survey is not necessarily what it was in the past. Also, in the IR file, v209, v238, and v208 give the number of births the respondent had in the past year, 3 years, and 5 years (years ago, not calendar years), respectively.

Subject: Number of respondents
Posted by Ritapriya Bandyopadhyay on Tue, 02 Jul 2024 08:43:05 GMT
View Forum Message <> Reply to Message

Hi,

I am trying to look at the number of women who were married before the age of 18 in Kenya - to understand child marriage. I saw in DHS 2014 report that 23% of women were married before the age of 18 - but there's also a column of 'number of respondents' - what does this mean? As in 23% of women aged 15-49 were married before the age of 18? If so, then the ,number of respondents, should be the total number of 15-49 year old women who were interviewed right? But it's much lower

Subject: Re: Number of respondents
Posted by Bridgette-DHS on Tue, 02 Jul 2024 11:56:22 GMT
View Forum Message <> Reply to Message

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

Users are asked to specify the table they are referring to. You are referring to table 4.3 in the Kenya 2014 report.

In this table, the first column gives the age range of the denominator. The percentage you give (22.9%) is in the row for women whose current age is 20-24. The weighted number of women age 20-24 in this survey is 5,375.

Subject: Re: Number of respondents
Posted by Ritapriya Bandyopadhyay on Wed, 03 Jul 2024 08:04:29 GMT
View Forum Message <> Reply to Message

Hi,
Thank you so much,

Another help: For 2022 Kenya DHS, I am trying to figure out the number of women aged 15-24 who got married before the age of 18 last year and women aged 15-19 who got pregnant last year. So the variables from the individual women's file that I will be using are the following right: v511, v508, v213, v214?

Could you also tell me how to use individual women weights here to calculate weighted estimates?

Best
Ritapriya

Subject: Re: Number of respondents
Posted by Bridgette-DHS on Wed, 03 Jul 2024 15:05:16 GMT
View Forum Message <> Reply to Message

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

Yes, you could produce such an estimate with those variables.

In Stata, you can see how to use weights by entering "help weight". There are many examples in forum responses and in the GitHub Stata programs. Also see the forum FAQ.

Subject: Re: Number of respondents
Posted by Ritapriya Bandyopadhyay on Thu, 04 Jul 2024 08:02:39 GMT
View Forum Message <> Reply to Message

Thanks so much,
Could you also help me understand the difference between current cohabitation (v509a) and first cohabitation (v509) - they are not same no? There's very less observations for v509a - that's why I was wondering.

Best
Reeth

Subject: Re: Number of respondents
Posted by Bridgette-DHS on Fri, 05 Jul 2024 19:16:35 GMT
View Forum Message <> Reply to Message

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

The current partner is not necessarily the first partner. Look at v503 and the relevant sequence of questions in the questionnaire. For some purposes, what you need is the date of first cohabitation, v509. DHS does not ask about all partners, but if there has been more than one, then the current partner is not the first partner, and v509a gives the date of cohabitation with that person.

Subject: Re: Number of respondents
Posted by Ritapriya Bandyopadhyay on Fri, 16 Aug 2024 13:11:52 GMT
View Forum Message <> Reply to Message

Hi thank you so much for the response,

Also, big THANKS for updating the education estimates for Kenya DHS 2022. I have a few very very basic questions:

(a) I understand that weighting is done to make the sample size representative of the population. However, for schooling data where we use household weight, let's say i type: tab schecyr_i [iweight =hhweight] - where schecyr_i is school enrollment and hhweight is v005/1000000 - since we use hhweight, does it mean that we are assuming that all members of the household are in school?

(b) If I want to compare counties based on school enrollment numbers - since we are making a comparison here - is it better to weight the percentage of children enrolled or the absolute numbers? SO let's say I want to look at the top 10 counties with the lowest school enrollment

Best
Reeth

Subject: Re: Number of respondents
Posted by Bridgette-DHS on Fri, 16 Aug 2024 16:26:01 GMT
View Forum Message <> Reply to Message

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

The weights compensate for the over-sampling of clusters in smaller strata and the under-sampling of clusters in larger strata, as well as for variations in the number of vacant households and non-response. That's why the weighted numbers are more representative than the unweighted numbers. All households within a cluster have the same weight. All individuals within a cluster have the same weight.

In (a), there is NOT an assumption that all household members are in school. Don't worry about that.

For (b), I would say there are 2 ways to rank the counties. Both of them use weights. You can rank them in terms of the number of students currently attending, or in terms of the percentage of eligible students who are currently attending. The following lines do this, using the PR file for the 2014 survey. I use the "collapse" command. There are alternatives, but this seems the easiest.

```
use "...KEPR72FL.DTA", clear

* measure of school attendance: sh18, "attend school current year"
* county list: shregion

gen cases=1
gen inschool=1 if sh18==1

collapse (sum) cases inschool [iweight=hv005/1000000], by(shregion)

* sort the counties by number currently attending
sort inschool
gen cases_rank=_n
list, table clean

* sort the counties by current attendance rates
gen inschool_pct=100*inschool/cases
sort inschool_pct
gen rate_rank=_n
list, table clean
```

Subject: Re: Number of respondents
Posted by Ritapriya Bandyopadhyay on Mon, 19 Aug 2024 14:42:48 GMT
View Forum Message <> Reply to Message

Hi
THANKS A LOT!
Two more things:

(a) Is it possible that the counties with lowest school enrollment, will differ depending on whether we are ranking them by absolute numbers or weighted percentages? If so, what are the implications of that?

(b) Secondly, in Kenya DHS 2022, I see that a lot of students aged 12-17 are enrolled in primary level, when I do tab hv122 agecat - agecat is age category - 6-11, 12-17. So a lot of students aged between 12-17 are in primary level - was wondering if there's any mistake here? - picture uploaded

Best
Ritapriya

File Attachments
1) Capture.PNG, downloaded 30 times

Subject: Re: Number of respondents
Posted by Bridgette-DHS on Mon, 19 Aug 2024 16:26:18 GMT
View Forum Message <> Reply to Message

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

Maybe I am misinterpreting what you mean by a percentage, but I am thinking that what you mean is the percentage of eligible children/people who are actually attending. If that's what you mean, then it is definitely possible for the number who are in school to be small and the percentage to be large, or for the number to be large and the percentage to be small. The ranking by number and the ranking by percentage can be very different. Weighting has nothing to do with this.

If someone age 12-17 is in primary school rather than secondary school, they could have started primary school late, or not have passed from one grade to the next. There could be a classification error. I looked at age in more detail with "tab hv105 hv122 if hv105<25". The correspondence between age and current level of school is not what I would expect but I can't explain why primary school seems to cover such a wide range of ages.