Subject: Issues with Honduran DHS dataset 2011-12
Posted by Mariela Contreras on Fri, 21 Feb 2020 03:03:06 GMT

Dear all,

We are having issues using the 2011-12 Honduran dataset.

When you see the flowchart (attached in file), one can see that n is 5,487 which means that the observations in the database are 5,627 minus the 140 women that have children with missing nutritional status values.

Using the command tab1 which is part of the epicalc package in R, one can see that the number of women living in the rural and urban areas is 3878 and 1749, respectively (attached in file). These numbers add up to 5,627 women which is n before excluding women who have children with missing nutritional status values. Using this command we have not adjusted for survey design.

When we generated a 2x2 table (using the variables place of residence (urban and rural) and child stunting category) with the survey package in R and adjusting for survey design, one can notice that n is 4,248 mothers with children and not 5,487 mothers with children as we have defined in our flow chart (attached in file).

We also noticed that when we use the svyby commands to generate prevalence, the results are difficult to interpret, as they are not prevalences but look more like integers (e.g. 1 and 2 for place of residence.

Please note the several commands we used to generate the above results:

1) To adjust for survey design:
dhsdesign <- svydesign(id= violnutr_3R$prisam, strata = violnutr_3R$stratasam, weights= violnutr_3R$samweight/1000000, data= violnutr_3R)

-prisam is v021
-Stratasam is v022
-Samweight is d005 (weight for domestic violence module)

2) To generate a frequency of the variable place of residence

tab1(violnutr_3R$plares)

-violnutr_3R is the name of the dataset
-plares is the variable name for the places of residence (urban/rural)

3) To generate a 2x2 table (place of residence and child stunting category)

svyby(~chstunting_cat,~plares,dhsdesign, svymean, na.rm=TRUE)

We would appreciate your support in sorting out our issue.

Best regards,

Mariela

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Bridgette-DHS on Fri, 28 Feb 2020 13:05:55 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I can't go through your code to confirm that it is doing what you want. I'm not an R user and besides that I'm not clear what it is you want to do. Are you trying to estimate the prevalence of stunting in urban areas and in rural areas, with adjustments for the survey design? I can do this very easily in Stata but I don't know the commands in R. The point estimates should be available in the report on this survey and on STATcompiler. You can compare with them. If the report and STATcompiler differ, the latter is preferred.

One issue with the stunting (etc.) estimates is that you can get them from the PR file (for all children in the household) or from the KR file (for those children whose mother is also in the household and is a de facto resident). If you are trying to match the report, you need to check whether the estimates you want to match come from the PR file or the KR file.

The point estimates are affected by the weights. The adjustments for clustering and stratification will only affect the standard errors of the estimates, i.e. the confidence intervals.

The weighted and unweighted frequencies are usually different for subpopulations. Sometimes the weighted frequencies are larger than the unweighted frequencies, and sometimes the reverse. That's not a problem. Tables in DHS reports give the weighted frequencies.

However, the weighted and unweighted frequencies should agree exactly for the TOTAL number of cases in the PR file and the IR file, because of how the weights are normalized.

Let me know if you have other questions--sorry I can't be of much help with this.

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Mariela Contreras on Mon, 23 Mar 2020 17:21:57 GMT
View Forum Message <> Reply to Message

Thank you for your reply Tom and Bridgette.

I am analyzing the association between intimate partner violence (IPV) and child stunting in Honduras. I am using the 2011 dataset.

I wonder why I get a smaller population size (weighted observations) than the number of observations (unweighted) during analysis? To keep track of what I have done so far, below my procedure in Stata.

Thank you for your support,

Mariela

 === ========
PROCEDURES
 === ========

I downloaded the dataset HNIR62FL.DTA and followed the steps.
1. Setting the sampling design features as:
****
*WEIGHT VARIABLE
gen weight = d005/1000000

*SURVEY SET
gen psu =   v021
gen strata = v022
svyset psu [pw = weight], strata(strata)
****
2. Keeping records that responded the domestic violence questions (keep if v044==1)
3. Generating the variable "**Women age 15-49 who have experienced physical violence since age 15".
****
gen everpsyvio=1 if (d105a>=1 & d105a<=4)|(d105b>=1 & d105b<=4)|(d105c>=1 & d105c<=4)|(d105d>=1 & d105d<=4)|(d105e>=1 & d105e<=4)|(d105f>=1 & d105f<=4)|(d105g>=1 & d105g<=4)|(d105j>=1 & d105j<=4)|(d130a>=1 & d130a<=4)
replace everpsyvio=1 if d115y==0
replace everpsyvio=1 if d118y==0
replace everpsyvio=0 if everpsyvio==.
****
4. Generating the variable corresponding to "*Persons Committing Physical Violence //Current husband/partner"
****
gen current=0 if everpsyvio==1
replace current=1 if v502==1 & ((d105a>=1 & d105a<=4)|(d105b>=1 & d105b<=4)|(d105c>=1 & d105c<=4)|(d105d>=1 & d105d<=4)|(d105e>=1 & d105e<=4)|(d105f>=1 & d105f<=4)|(d105j>=1 &

```
    d105j<=4))
    replace current=1 if v502==1 & d118a==1
    ********************************************************
```

5. Generating the variable child stunting. I use the last alive child with height measurements.
```
    ********************************************************
```

**the youngest child's height
```
keep if  hw70_1 < 9996
```

**computing height for age < -2
```
gen haz=hw70_1/100 //converting to meters
gen pstunted=0
replace pstunted=1 if haz<-2
replace pstunted=. if haz==.
tab pstunted
********************************************************
```

6. Crosstable between physical violence by current husband/partner and child stunting

tab pstunted current //unweithed

```
     |       current
 pstunted |      0        1 |    Total
-----------+---------------------+----------
       0 |     800      611 |    1,411
       1 |     221      245 |     466
-----------+---------------------+----------
   Total |    1,021      856 |   1,877
```

svy: tab pstunted current, col //weighthed

(running tabulate on estimation sample)

```
Number of strata  =       38            Number of obs    =     1,877
Number of PSUs    =      892            Population size  = 1,635.6582
                              Design df       =      854
```

```
-------------------------------
         |      current
 pstunted |    0     1  Total
----------+--------------------
       0 | .8063 .7385 .7765
       1 | .1937 .2615 .2235
         |
   Total |    1     1     1
-------------------------------
  Key:  column proportion
```

```
  Pearson:
    Uncorrected   chi2(1)      =  12.2383
```

Design-based F(1, 854)   =   8.7656   P = 0.0032

.
end of do-file

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Bridgette-DHS on Mon, 30 Mar 2020 19:59:19 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I can't identify, in what you sent, what are the weighted and unweighted frequencies. However, I think you have done this correctly, using the DV weight (d005) and restricting to the children of women who were administered the DV module (v044==1).

Weighted and unweighted frequencies usually differ but are within about 10% or so of each other. They will only agree exactly for the overall total number of cases. Please tell me if you are seeing a discrepancy bigger than 10% or so.

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Mariela Contreras on Tue, 31 Mar 2020 02:17:52 GMT
View Forum Message <> Reply to Message

Thank you for your message Tom and Bridgette.

Indeed, we are seeing a discrepancy bigger that 10% between weighted and unweighted frequencies. In the analyses we sent you all, the weighted frequency was 1635 observations and the unweighted frequency was 1877 observations.

Best regards,
Mariela

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Bridgette-DHS on Tue, 31 Mar 2020 13:32:27 GMT
View Forum Message <> Reply to Message

Following is another response from DHS Research & Data Analysis Director, Tom Pullum:

Your weighted total is 13% less than the unweighted total. That's plausible. I looked at the file and it appears that the weighted total is indeed about that much less than the unweighted total, although I have different counts. This happens when fertility is different in the under-sampled areas than it is in the over-sampled areas. You then get a bigger discrepancy for the child file than for the woman file.

I believe you are trying to relate stunting to domestic violence. Right? There's an easier way to construct your data file. After opening the IR file, run these Stata lines:

keep v001 v002 v003 v005 d005 hw70_*
keep if d005<.

reshape long hw70_, i(v001 v002 v003) j(bidx)
rename *_ *
drop if hw70==.
gen stunted=0 if hw70<600
replace stunted=1 if hw70<-200

You will want many more variables in the "keep" line, including the stratification variable, but these lines will give you a child file for all the children whose mother was in the same household and was given the DV module.

Subject: Re: Issues with Honduran DHS dataset 2011-12
Posted by Mariela Contreras on Wed, 01 Apr 2020 00:30:19 GMT
View Forum Message <> Reply to Message

Thank you very much for your support, Tom and Bridgette. Indeed, we are studying the relationship between domestic violence and child stunting, including other forms of malnutrition. We will go ahead and continue with the analyses.

Best regards,

Mariela