
Subject: IWeight -EDHS

Posted by [reemnabil](#) on Mon, 05 Mar 2018 16:18:17 GMT

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

Hi,

I am using the 1995, 2003 and 2014 EDHS - Egypt child Recode datasets as I want to look at the changes in stunting over time. I am doing an oaxaca and dfl decomposition. I plotted the density of the Height - for - age Z score (HAZ), once using weights and once without weights and the graphs look different. I used the following code for weights

```
gen sampwt = v005/1000000
```

```
gen psu = v021
```

```
gen strata = v022
```

```
tab stunting [iweight=sampwt]
```

```
svyset[pw=sampwt], psu(psu) strata(strata)
```

Now, when I run the dfl code I can't use weights, and so I know that the result I get will not be a nationally representative one. I watched the DHS video and it said to use iweights. My question is why should we use iweights instead of fweights, and if I want to run the dfl, and it doesn't allow for weights, is there a way I can change the variables to include the weights and then run the dfl on the weighted data.

Thank you,
Reem

Subject: Re: IWeight -EDHS

Posted by [Bridgette-DHS](#) on Thu, 08 Mar 2018 14:32:53 GMT

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

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

First a response to the question "why should we use iweights instead of fweights?" The correction for the sample design with svyset uses pweight--not iweight or fweight. The weights are proportional to the inverse of the sampling probability and are used to produce unbiased estimates. pweight does the appropriate adjustment for over- and under-sampling of strata. Some Stata commands do not accept the pweight option but do accept iweight. Just speaking for myself, that's the only reason why I would ever use iweight, as kind of a trick for Stata to get it to do what I want for commands that don't accept pweight. I believe the adjustments always match with pweight, when you could use either, but I have less faith in iweight. fweight is very different. It's analogous to the "expand" command in Stata. Although the calculations are not the same, its effect is equivalent to producing extra observations. Here's a simple example:

```
set obs 2
```

```
gen x=_n
```

```
summarize x
```

```
gen n=10
summarize x [fweight=n]
expand n
summarize x
```

If you paste these lines into the Stata command line, you will see exactly the same result for the second and third "summarize" commands. Sometimes I will trick Stata by using fweight. However, when doing that, you must correct the standard errors. For example, if you artificially increase the sample size by a factor of 10, then you will artificially reduce the standard deviation by a factor of $1/\sqrt{10}$. fweights must be integers. v005 has a factor of 1000000 and no decimal place, so if you use it as an integer, you will get unbiased estimates but the standard errors will have to be multiplied by 1/1000. You could simulate using fweights with "expand v005" but the number of "cases" would then be unmanageable. Another option with Stata is aweight and it is almost never appropriate. If you want to use it, be sure to read about it.

Finally I can just say a little about your main question, whether to use the decomposition procedure. Within DHS, when we want to use a procedure that does not allow weights, we have to decide which is more important--the procedure or the adjustments for sample design. I recommend that you use the decomposition procedure and see whether it is helpful for the interpretation of the data. If you report the results, you will have to include caveats about bias in the estimates and the questionable standard errors. We can all hope that such procedures will eventually appear in software packages with the adjustments for the sample design.
