

---

Subject: Pooled Cross sections

Posted by [cbdolan](#) on Tue, 14 Jun 2016 13:35:27 GMT

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

---

I am using the 2007 and 2013/14 DRC Birth Recode files.

Using DHS forum guidance, I have de-normalized the weights using process outlined below. Then I appended the 2007 and 2013/14 files to start the analytic data set. I have the following specific questions:

1. To get the sample size of women 15-49 I used the individual recode. Is this typical or is there another file/source researchers use for that N?
2. There is a debate on what to do with these de-normalized weights and the use of the stata sub-pop command. I don't know if I can just use the weights as they are (de-normalized) or if I have to do something else before I can use the subpop command at the cluster level (ie. running rural and urban sub-pop analysis).
3. I have been told that if I am using individuals as the units of observation in a regression, then we don't use the sampling weights at all. The thought is that other individual level controls should likely pick up any of the differences in weights. Is this assumption correct?
4. Additionally, some researchers are more comfortable using using microeconomic methods (appropriately clustering standard errors) instead of these survey weights, particularly when using data from more than one survey. Have you come across any references in the literature (or even forum posts) that discusses these advantages/disadvantages with the DHS data?

Thanks in advance for your time.

```
*I'm adding in code to use for pooled weights
use "Y:\Data\4_DHS_BirthRecode\CDBR61FL.dta"
*Original weight in DHS : v005 (which should preferably be divided by 1000000)
generate n_v005=(v005/1000000)
*note this is the population of 15-49 in DRC (2013) from United Nations, Department of Economic
and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision,
custom data acquired via website.
generate P1549=16167000
*note this is the sample size from the individual recode file of women 15-49 interviewed
generate n1549=18827
*Country specific weight :CSW= P1549/n1549 (population aged 15-49 in the country / sample size
of )
generate CSW=(P1549/n1549)
*New weight
generate NW=n_v005*CSW
file Y:\Data\4_DHS_BirthRecode\n_CDBR61FL.dta saved
clear
```

```
use "Y:\Data\4_DHS_BirthRecode\CDBR50FL.dta"
*Original weight in DHS : v005 (which should preferably be divided by 1000000)
generate n_v005=(v005/1000000)
*note this is the population of 15-49 in DRC (2013) from United Nations, Department of Economic
and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision,
custom data acquired via website.
generate P1549=13201000
*note this is the sample size from the individual recode file of women 15-49 interviewed
generate n1549=9995
*Country specific weight :CSW= P1549/n1549 (population aged 15-49 in the country / sample size
of )
generate CSW=(P1549/n1549)
*New weight
generate NW=n_v005*CSW
file Y:\Data\4_DHS_BirthRecode\n_CDBR50FL.dta saved
clear

*generate weight: see code at top
*make unique strata values by region/urban-rural )
egen stratum=group(v024 v025)
*tell stata the weight (using pweights for robust standard errors, cluster (psu), and strata
svyset [pw=NW],psu(v021)strata(stratum)
*prefix regrss with "svy:stata will now know how to weight your data and compute the right
standard errors */
```

---