Subject: urban-rural variable
Posted by menonidhi on Tue, 21 Apr 2015 14:14:44 GMT
View Forum Message <> Reply to Message

In my study, I have two variables - V025: which classifies place of residence as Urban/Rural and S025: which classifies residence as Mega city/Large City/Small City/Large Town/Small town/Rural.

I was able to extract the former only for females. To check if I could could substitute V025 with S025(which I was able to extract for both the sexes), I checked the distribution of urban and rural population among females. I clubbed the values of Mega city/Large City/Small City/Large Town/Small town as Urban.

I however observe few discrepancies. The number of females in the urban area is 18,442 and rural is 20,815. In the second variable, the number of females in rural area is 23,529. I have summed the values of mega city,large city,small city, large town and small town which comes to 15,728.
I observe an exact difference of 2,714 i.e. (23,529-20,815=2714) and (18,442-15,728 =2,714). I have attached my results for any reference for the values mentioned above.

Is there a reason for this difference.

File Attachments
1) urban-rural.txt, downloaded 690 times

Subject: Re: urban-rural variable
Posted by Trevor-DHS on Fri, 24 Apr 2015 19:21:50 GMT
View Forum Message <> Reply to Message

It wasn't very clear from your explanation which survey and datasets you are using, but it appears that you are using the HIV tests results data, matched with other datasets. However it appears that there is a problem with the matching of your data. I would expect the number of women and men to be roughly equal, but you have twice as many men as women. Below is a short piece of code for matching with the persons recode file, and for checking the sex distribution:

```
cd "C:\Data\DHS_Stata"
use "IAPR52FL.dta"
clonevar hivclust = hv001
clonevar hivnumb = hv002
clonevar hivline = hvidx
merge 1:1 hivclust hivnumb hivline using "IAAR51FL.dta"
keep if _merge==3
tab hv104
tab hv025
tab sh025
tab sh025 hv025
```

From this, I received the following output:

```
. tab hv104
```

```
    sex of |
  household |
    member |      Freq.     Percent        Cum.
-----------+-----------------------------------
      male |     51,641       48.88       48.88
    female |     54,016       51.12      100.00
-----------+-----------------------------------
     Total |    105,657      100.00
```

```
. tab hv025

   type of |
   place of |
  residence |      Freq.     Percent        Cum.
------------+-----------------------------------
      urban |     52,697       49.88       49.88
      rural |     52,960       50.12      100.00
------------+-----------------------------------
      Total |    105,657      100.00
```

```
. tab sh025

city\town\c |
 ountryside |      Freq.     Percent        Cum.
------------+-----------------------------------
  mega city |      8,696        8.23        8.23
 large city |     21,989       20.81       29.04
 small city |      6,352        6.01       35.05
 large town |      1,671        1.58       36.64
 small town |     13,989       13.24       49.88
      rural |     52,960       50.12      100.00
------------+-----------------------------------
      Total |    105,657      100.00
```

```
. tab sh025 hv025

city\town\ |   type of place of
countrysid |       residence
         e |     urban      rural |     Total
-----------+----------------------+----------
 mega city |     8,696          0 |     8,696
large city |    21,989          0 |    21,989
small city |     6,352          0 |     6,352
large town |     1,671          0 |     1,671
small town |    13,989          0 |    13,989
     rural |         0     52,960 |    52,960
-----------+----------------------+----------
```

Total | 52,697 52,960 | 105,657

As you can see in my tables (unweighted) there are about 49% men and 51% women, which is roughly what I would expect.

Also, as you can see the recoding of urban and rural makes sense in the last table.
