
Subject: merging WI with IR/MR for Zambia (code for South Africa will not work)

Posted by [alitasso](#) on Thu, 09 Jan 2014 14:32:05 GMT

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

I am trying to merge the WI file with the IR/MR files for Zambia 2001-02. In this post (in case the URL dies, the post is titled, "merging wealth index from Ethiopia DHS 2000 to women's file"), Tom Pullum provides some code to merge the WI file with the IR file for Ethiopia 2000. However, it does not seem to work for Zambia.

First I look at the range of v001 and v002 values in the IR file:

```
. use ZMIR42FL.dta, clear
```

```
. tabstat v001 v002, stat(min max)
```

```
stats |   v001   v002
-----+-----
min |     1   1001
max |   320   9469
-----
```

Then I try to apply Tom Pullum's code to the WI file:

```
. use ZMWI41FL.dta, clear
```

```
. gen str9 wv001=substr(whhid,1,9)
```

```
. gen str9 wv002=substr(whhid,10,3)
```

```
. destring wv001, gen(v001)
```

wv001 has all characters numeric; v001 generated as int

```
. destring wv002, gen(v002)
```

wv002 has all characters numeric; v002 generated as int

```
. tabstat v001 v002, stat(min max)
```

```
stats |   v001   v002
-----+-----
min |    15     1
max |  3201   469
-----
```

As you can see, the range of v001 and v002 codes generated by Tom Pullum's code does not match the range of v001 and v002 codes that are in the IR data. It looks like it's just off by just one column: v001 should be 3 digits, and v002 should be 4 digits. But it doesn't seem to be that simple.

If I modify the code slightly, then I get some unwanted spaces in v002:

```
. use ZMWI41FL.dta, clear  
  
. gen str9 wv001=substr(whhid,1,8)  
  
. gen str9 wv002=substr(whhid,9,4)  
  
. destring wv001, gen(v001)  
wv001 has all characters numeric; v001 generated as int  
  
. destring wv002, gen(v002)  
wv002_ contains nonnumeric characters; no generate
```

However, this code seems to work fine for v001:

```
. tabstat v001, stat(min max)
```

variable	min	max
v001	1	320

I tried a crude solution of just deleting the spaces in wv002, but then I get a variable that does not range from 1 to 9469 (as it should):

```
. gen str9 wv002_=substr(wv002," ","",.)  
  
. destring wv002_, gen(v002_)  
wv002_ has all characters numeric; v002_ generated as int  
  
. tabstat v002_, stat(min max)
```

variable	min	max
v002_	11	9469

Any assistance would be appreciated -- thank you!
