Subject: Merging Women-Men-HIV Data: Different Countries, Different Years
Posted by Yawo on Sun, 10 Mar 2019 21:24:46 GMT
View Forum Message <> Reply to Message

Hello,

I am getting started on big project investigating stigmatization among those living with HIV in sub-Saharan Africa. Given the relatively low rates of testing in SSA, I needed to pool data across countries to ensure my analyses (HLM) models have enough power.

Consequently, I am pooling data from 33 countries with HIV data: Angola Benin Burkina Faso Burundi Cameroon Chad Congo Congo Democratic Republic Cote d'Ivoire Ethiopia Gabon Gambia Ghana Guinea Kenya Lesotho Liberia Malawi Mali Mozambique Namibia Niger Rwanda Senegal Sierra Leone South Africa Swaziland Tanzania Togo Uganda Zambia Zimbabwe

For some countries, like Kenya, we have HIV data for 2 years: 2003 and 2008, while for others like Zimbabwe, we have data for 2005, 2010 and 2015

So, for each country, I need to:

1. append men and women's data, then merge this with HIV test results for a specific year;
2. repeat the same for the next year,
3. append data of year1, year2, year3;
4. Then finally, pool all these data together to create a dataset for all countries, for all years for which they HIV data.

I have already:
a) selected a subset of variables I needed to be sure they are consistent across countries, and years; including the various survey-specific variables - stratum, psu, etc
(b) renamed the male variables, from mv* to v*;
(c) dumped all the datasets (men, women, HIV) into one main datafile, that I will use as my working directory.

Here is the process I have outlined, and I will appreciate some comments and suggestions:

/* sort Kenya2003 women's data by key variables */

```
use Kenya2003_Individual.data, clear
sort v001 v002 v003
save Kenya2003_women1.dta, replace
```

/* sort Kenya2003 men's data by key variables - note, variables already renamed from mv* to v* */

```
use Kenya2003_male.dta, clear
sort v001 v002 v003
```

```
save Kenya2003_male1.dta, replace

/* call Kenya2003 HIV data, rename key variables, then sort */
use Kenya2003_HIV.dta
ren hivclust v001
ren hivnumb v002
ren hivline v003
sort v001 v002 v003
save Kenya2003_HIV1.dta, replace

/* Append Kenya2003 men to women */
use Kenya2003_women1.dta, clear
append using Kenya2003_male1.dta
save KenyaWomenMen.dta, replace

/* Merge HIV data into the combined Kenya 2003 men and women file */
use KenyaWomenMen.dta, clear
sort v001 v002 v003
merge merge v001 v002 v003 using Kenya2003_HIV1.dta
save Kenya2003_HIV_MenWomen.data, replace.

/* Repeat same steps to create Kenya2008_HIV_MenWomen.data */
use Kenya2003_HIV_MenWomen.dta
append using Kenya2008_HIV_MenWomen.dta
save Kenya2003-2008_MenWomenHIV.dta, replace
```

Cycle the same process through all countries. At the end, append, country datasets to each other.

Questions:

1. Assuming everything is correct, how do I handle the psu, stratum, etc variables needed to create my svyset? Do I do this for each dataset (within a country, within a specific year) or wait until the end, and create a grouping variable to do this?

2. Any other advice to help / facilitate this process?

Thanks very much in advance for any assistance -

best- Yy