

Dear DHS Team,

I am conducting research on "The Effect of Girls' Empowerment on Adolescent Pregnancy in Sub-Saharan Africa," aiming to investigate whether increased aged 15-19 girls' empowerment has a positive effect on reducing adolescent pregnancy rates in this region.

I plan to pool data from 27 Sub-Saharan African countries and will be using DHS-7 and DHS-8 data from the IR datasets of these countries. The explanatory variable will be women's empowerment, while the dependent variable will be the pregnancy status of adolescents aged 15-19. I intend to perform logistic regression analysis using Stata. I have used the "append" function to pool the data from the 27 countries into one dataset.

I have been using the SWPER Global Index by Ewerling et al. (2020) as a tool to measure women's empowerment. I have attached the relevant Stata do-file for your reference.

When I run the SWPER Global Index code using data from a single country, I encounter no issues. However, when I pool data from 27 countries and then attempt to run the code, I experience several errors.

I am not sure if this question is appropriate for this forum, but I thought I would ask in case you could provide any guidance.

The errors occur in the section of the SWPER Global Index code titled //Wm autonomy questions, specifically during the execution of the section labeled *Imputing age1birth for those women that do not have children***.

I have outlined the specific portion of the code where the errors occur below.

```
//Wm autonomy questions
clonevar age1cohab=v511
*Imputing age1birth for those women that do not have children***
recode age1cohab 33/max=33, gen (age1)
hotdeck v212, store by(age1) keep(caseid) imp(1)
sort age1 v212
preserve
use "imp1.dta", clear
rename v212 v212_i
drop age1
save, replace
restore
cap drop _merge
merge 1:1 caseid using "imp1.dta"    --> variable caseid does not uniquely identify observations
in the master data r(459);
erase "imp1.dta"
clonevar age1birth=v212_i    --> variable v212_i not found  r(111);
```

merge 1:1 caseid using "imp1.dta" --> variable caseid does not uniquely identify observations in the master data r(459);

To address this issue, I executed the following command:

```
.duplicates drop caseid, force
```

This resulted in the deletion of 187 observations, as shown below:

Duplicates in terms of caseid (187 observations deleted)

I was wondering if there might be an alternative solution. Since 187 observations are deleted with this method, I would prefer another approach if possible.

However, even after resolving duplicates in this way, the following error still occurs:

If you have any suggestions on how to resolve these issues, I would greatly appreciate your help.

File Attachments

1) [SWPER_global \(1\).do](#), downloaded 152 times

Subject: Re: Issues with Pooled Multi-Country DHS Data
Posted by [Bridgette-DHS](#) on Fri, 08 Nov 2024 17:07:23 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS staff member, Tom Pullum:

This question would take more time than DHS staff can allocate to the forum. A strategy to deal with the problem would be to loop through the 27 surveys, applying the model to every one of them individually. You will probably be able to localize the problem into one (or a few?) surveys.

For some reason you are doing some imputation. I also suggest removing that step or bypassing it for testing purposes.

Subject: Re: Issues with Pooled Multi-Country DHS Data
Posted by [schoumaker](#) on Sat, 09 Nov 2024 09:09:46 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

If you append data, you should make sure each case has a single identifier. I see you use caseid in your code, and you have duplicates for caseid - it may be part of the problem. Could you also

explain what your imp1.dta is? file is? As Tom suggested, you may remove the imputation section. I also do not see why you would impute age at first birth for women who have not had a birth; since they have not had their children, imputing their age at first birth may not be a relevant approach. Since you plan to use a logistic regression on the pregnancy status of adolescents, imputing age at first birth does not seem necessary.

Hope this helps.

Bruno
