
Subject: Issues with Pooled Multi-Country DHS Data
Posted by [ygkim127](#) on Fri, 08 Nov 2024 15:04:32 GMT
[View Forum Message](#) <> [Reply to Message](#)

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 14 times
