
Subject: Egypt DHS 2008 - Merging EGIR5AFL and EGOD5AFL

Posted by [mariacedro](#) on Tue, 22 Mar 2022 15:06:04 GMT

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

Hello,

I am working with the 2008 Egypt DHS data, and I have to merge the dataset containing questions about AIDS and HIV (this is dataset EGOD5AFL) to EGIR5AFL, which contains information on circumcision, fertility and family planning. This is the STATA code I am using:

```
use "/Users/maria/Dropbox/Egypt DHS/Raw Data/Egypt  
DHS/2008/EGOD5ADT/EGOD5AFL.DTA", clear
```

```
rename hpsu v001  
ren hnumber v002  
ren wline v003
```

```
sort v001 v002 v003
```

```
merge 1:1 v001 v002 v003 using "/Users/maria/Dropbox/Egypt DHS/Raw Data/Egypt  
DHS/2008/EGIR5ADT/EGIR5AFL.DTA", force  
tab _merge
```

However, using this code I get a pretty low matching rate, only equal to 17%.

Is there something I am getting wrong?

Thank you.

Best,

Maria

Subject: Re: Egypt DHS 2008 - Merging EGIR5AFL and EGOD5AFL

Posted by [Bridgette-DHS](#) on Wed, 23 Mar 2022 11:33:51 GMT

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

Following is response from DHS Research & Data Analysis Director, Tom Pullum:

On page xxiii of the report you will find this paragraph: "The 2008 EDHS also collected information on a number of other health topics from 6,578 women and 5,430 men age 15-59 living in a subsample of one in four of the households surveyed. Among the key topics covered in these interviews were knowledge and awareness of avian influenza, HIV/AIDS and hepatitis C; previous history of hypertension, cardiovascular illness diabetes and liver disease; attitudes and behavior with respect to female circumcision; health care costs; and health insurance coverage."

I did the merge with the lines given below. I reduced the respondents in the OD file to ever-married women age 15-49. Then, out of the 16,527 women in the IR file, I was able to find 4,181, almost exactly one in four, in the OD file. This is consistent with one fourth of the households being selected for the OD survey.

You merged with about 17% of women, rather than 25%, because you retained men, women age 50-59, and never-married women in the OD file. However, your merge was ok.

When users find a result like this, I recommend checking for subsampling on some of the key variables. Just go to the pdf of the final report and search for the word "subsample" and you will probably find the explanation.

```
cd e:\DHS\DHS_data\scratch
```

```
use "...EGIR5AFL.DTA", clear
keep v0*
gen cluster=v001
gen hh=v002
gen line=v003
gen in_IR=1
sort cluster hh line
save EGtemp.dta, replace
```

```
use "...EGOD5AFL.DTA", clear
keep wline h* i*
```

```
* restrict to women
keep if h009==2
```

```
* restrict to ever-married
keep if i112<6
```

```
* restrict to age 15-49
keep if i111>=15 & i111<=59
```

```
gen cluster=hpsu
gen hh=hnumber
gen line=wline
gen in_OD=1
sort cluster hh line
merge cluster hh line using EGtemp.dta
tab _merge
```

```
replace in_IR=0 if in_IR==.
replace in_OD=0 if in_OD==.
tab in_*
```

Subject: Re: Egypt DHS 2008 - Merging EGIR5AFL and EGOD5AFL
Posted by [mariacedro](#) on Wed, 23 Mar 2022 13:36:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

This was very helpful.

Thanks a lot!

Best,

Maria
