Hello everyone,

I hope you're all doing well.

I'm reaching out to request some support. I'm working with a database that has 40,394 observations. However, upon examining the report, I noticed that the total number of households is 15,705, from which information was obtained for 15,254 women of reproductive age, as mentioned in the report. Thus, I've come to understand that there are duplications in the database I possess and realized the need to remove them to reach the expected 15,254 observations.

However, my concern arose when I noticed that, upon attempting to remove the duplications to reach the expected number of surveyed women of reproductive age (15,254), the total observations drastically decreased to 257 when using the V002 variable, while the V003 variable showed 30 observations. When employing the CASEID variable, I obtained 11,116 observations.

I would like to better understand how I can overcome this issue and obtain an accurate count of the surveyed women of reproductive age.

Thank you in advance for your attention, and I hope to receive guidance on how to proceed regarding this matter.

Best regards,

Sancho Xavier

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

First, you did not identify the survey you are working with, but I believe it is the Tanzania 2022 DHS (which is not an MIS). 15,705 households were selected. These are the cases in the HR file. The number of people listed in the households, as either de jure or de facto residents, is 40,394. These are the cases in the PR file. They include men, women, children. Eligible women for the interviews are women age 15-49 who are de facto residents. Some households have no such women, some have 1, some have 2, etc. The total number of women who were eligible AND were interviewed is 15,254. These are the cases in the IR file. There is an average of about one woman per household, but the ratio of women to households does not have to be this close to 1 and it certainly does not need to be exactly 1. The cluster ID is v001. Households are numbered within clusters, with v002 as the household ID. Individuals within the household are given a line number, and that's v003. "caseid" is a string variable that includes columns for v001, v002, and v003. It

takes 15,254 unique values.

If you open the IR file in Stata and enter "codebook v001 v002 v003 caseid" and "list v001 v002 v003 caseid if _n<=100, table clean" you will get a list of the ID variables for the first 100 women in the file. It may help you to see how the data are structured. As I said, there is no problem.

Subject: Re: DHS-MIS
Posted by Sancho on Fri, 01 Mar 2024 16:39:09 GMT
View Forum Message <> Reply to Message

I'm working with the TZBR82FL database from Tanzania (2022), and I'm analyzing the data using R. I would like to know what is necessary to obtain the mentioned number of 15,254 women. Should I request another file, or can I continue with the 40,394? This is my first experience working with DHS databases. I already have the TZBR82FL database for SPSS, which contains 40,394 observations.

File Attachments
1) Capturar.PNG, downloaded 23 times
2) Capturar1.PNG, downloaded 22 times

Subject: Re: DHS-MIS
Posted by Bridgette-DHS on Fri, 01 Mar 2024 18:12:17 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

There are different files for different units of analysis. The BR file ("BR" is in "TZBR82FL") is a file of births (the "B" is for "births"). Each case is a birth in the birth histories of women in the IR file (the "I" is for "individuals"). That file has 15,254 cases in it. I suggest that you switch to that file.

Please read more about dataset types here: https://dhsprogram.com/data/Dataset-Types.cfm

---

Subject: Re: DHS-MIS
Posted by Sancho on Fri, 01 Mar 2024 23:27:40 GMT
View Forum Message <> Reply to Message

I am writing to seek your guidance on which database you would recommend for studying anemia in children aged 6 to 59 months. I am aware that this outcome of interest is present in the TZBR82FL database. Considering my specific interest, I would like to know which database you suggest I use for my analysis.

It seems that I can use the PR file to obtain the prevalence of anemia in children aged 6-59 months and its associated factors, according to this guideline I read at this link: https://dhsprogram.com/data/Guide-to-DHS-Statistics/Anemia_S tatus.htm.

Subject: Re: DHS-MIS
Posted by Bridgette-DHS on Mon, 04 Mar 2024 16:03:45 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

You have a couple of earlier postings but this is the first time you have actually said what you are studying. The anemia data for children are in the BR and KR files (hw57) and in the PR file (hc57). You probably want to use the KR file. It includes other information about the child and the mother, as well as household-level variables. DHS Methodological Report 18 (https://www.dhsprogram.com/pubs/pdf/MR18/MR18.pdf) may be helpful.

Subject: Re: DHS-MIS
Posted by Sancho on Mon, 04 Mar 2024 16:25:16 GMT
View Forum Message <> Reply to Message

After analyzing the data available in the BR file, I was able to identify the association I was investigating, specifically the relationship between maternal anemia and the likelihood of a child also developing anemia. Based on these findings, I would like to confirm if I can proceed with just this file (BR), considering that the relevant data is contained within it and that I have already been able to answer my research question using only the data from the BR file. My main objective is to understand the association between maternal anemia and the likelihood of anemia in children.

Subject: Re: DHS-MIS
Posted by Bridgette-DHS on Mon, 04 Mar 2024 20:30:23 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

Yes, you can stay with the BR file, or you can switch to the KR file without losing anything. The KR file is like the BR file, with children as cases. But it is reduced to children born in the past 5 years, who are the only children who have anemia data. You would look at the relationship between v457 (for the mother) and hw57 (for the child).