Subject: Data name and data do not match
Posted by JaneQuan on Sun, 20 Jun 2021 13:57:22 GMT
View Forum Message <> Reply to Message

Hi,

Please check the attachment first. Thank you.
I downloaded 25 Sub-Saharan countries' data of different years (40 compressed files in total), but I find that 15 files' names(columns P, Q, R together are the name of the files) and the actual data(column S) don't match.

May I ask why?

Thank you:)

File Attachments
1) p.PNG, downloaded 284 times

Subject: Re: Data name and data do not match
Posted by Bridgette-DHS on Mon, 21 Jun 2021 13:23:01 GMT
View Forum Message <> Reply to Message

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

The 5th character in the file name is the phase of the DHS coding manual and the 6th character is the "version" if there are updates. For example, when the files are first issued, the 5th-6th characters for a Phase 6 survey could have been "60". If there was some correction or update, the files were re-issued with "61". Next time, "62". Usually, ALL the files from a survey (HR, PR, IR, MR, CR, BR, KR) will have their file names updated, even if the update does not affect the specific file. Sometimes (more in the past than in the present), if a file is not affected, it will not be renamed. In the examples you give, the MR file must not have been affected, so that file was not re-issued or re-named.

This is annoying for all of us who use the data, but when it happens, you have to adapt. For example, if you have a combination of IR62 and MR61, you just have to treat the MR61 file as if it were labelled MR62. The two files come from the same survey.

When there are two surveys in the same phase if DHS, the 6th character will begin as a letter, and then any updates will be the next letter. For example (these may not be the actual letters that would be used) the first survey may be in a sequence 60, 61, 62, as described above, but a second survey would be 6H, 6I, 6J... Then if there is a third survey within the same phase of DHS, the sequence will be 6R, 6S, 6T...

This numbering system leaves much to be desired. It goes back to decades ago, when file names were restricted to 8 characters. Eventually, I am sure, a more flexible and intuitive naming system will take over.

Subject: Re: Data name and data do not match
Posted by JaneQuan on Tue, 22 Jun 2021 01:25:09 GMT
View Forum Message <> Reply to Message

Hi,
First of all, thank you so much for explaining the data (which is the issue about Column V in the attachment).
And Yes, I have relabelled MR61 as MR62 and also other files when I combine them.

But what I really wanted to ake is why the data's filename does not match the actual data in the dta.file.
For example, Ethiopia (Row1), the filename is "Ethiopia: Standard DHS, 2011", but after I downloaded it and open the data file in the STATA, you will find the survey year is 2003.
Also, Burundi should only contain a one-year dataset if based on the filename, but the actual data contains a two-year dataset.

Another example, the Chad(Row 8), the survey phase in My Dataset Account is presented as VII (also can tell from the extracted filenames), but when you open the data in the STATA, the survey phase showed as TD6.

Thank you^^

Subject: Re: Data name and data do not match
Posted by Bridgette-DHS on Tue, 22 Jun 2021 13:06:21 GMT
View Forum Message <> Reply to Message

Ethiopia: All standard date variables in the Ethiopia data file are in the Ethiopian calendar - see this post.

Burundi: The Burundi DHS 2010 survey was collected August 2010 - January 2011.

Chad: The survey was collected under Phase VII but the recode structure (programs) used in the data is DHS VI (which explains "TD6").

Subject: Re: Data name and data do not match
Posted by JaneQuan on Wed, 30 Jun 2021 04:18:16 GMT
View Forum Message <> Reply to Message

Hi Bridgette,

Thank you so much for your answer.
You said "Burundi: The Burundi DHS 2010 survey was collected August 2010 - January 2011." so that's why the actual data includes both years of 2010 and 2011 while the filename is Burundi Standard DHS 2010.

I have two questions here (pls. check the attachment first):
1. Should I turn the data of two years into one year, such as changing 2011 to 2010? (maybe also

--Because the number of observations in two years has a huge gap.

2. Those filenames contain two years instead of one, such as Angola Standard DHS 2015-2016, is this a two-year time-series data instead of collected in two years?

Thank you.

File Attachments
1) capture.PNG, downloaded 285 times

Subject: Re: Data name and data do not match
Posted by Bridgette-DHS on Wed, 30 Jun 2021 11:57:28 GMT
View Forum Message <> Reply to Message

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

DHS surveys are household-to-household surveys with teams of interviewers in sampled geographic areas, working over a period of several months. Sometimes the fieldwork all takes place within one calendar year. Sometimes, as with this survey, it straddles two consecutive years. "2010" or "2010-11" is just part of the label for the survey. You definitely do not need to separate the survey into two parts for the two calendar years. If, say, the data came from a birth registration system, and births are reported by calendar year, then you would want to separate the two years, but that would not be appropriate for survey data.

Subject: Re: Data name and data do not match
Posted by JaneQuan on Wed, 30 Jun 2021 14:30:40 GMT
View Forum Message <> Reply to Message

Thank you for your reply, but I am still confused.

I quote from your previous answer"You definitely do not need to separate the survey into two parts

for the two calendar years. "
-my response is that I didn't separate the data, it is the original data includes both years while it should've only had one year of data according to the filename.
and I totally understand that "sometimes, as with this survey, it straddles two consecutive years".

but the problem for me is that "year" is one of the variables in my analysis model and since the sample size of the two consecutive years is quite different, maybe I should combine those data into one year as the filename implies.

Regards.

PS. This confuses me is because the year on other filenames and actual data are consistent, except for only 5 countries' among 40.

Subject: Re: Data name and data do not match
Posted by Bridgette-DHS on Wed, 30 Jun 2021 15:13:12 GMT
View Forum Message <> Reply to Message

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

Ok--I didn't realize that was the issue. I agree with selecting one of the years--the one in which there were more interviews. That is, use all the data but assign it to that year for a reference year.

Subject: Re: Data name and data do not match
Posted by JaneQuan on Thu, 01 Jul 2021 06:02:59 GMT
View Forum Message <> Reply to Message

Hi Dr. Pullum,

Thank you so much for the answer. So I should just select one year of data that relatively has more interviewers if the survey is conducted in two consecutive years (or the year of fieldwork spanned both years).

and I have one more question to ask on this thread: how often do you conduct a Standard DHS survey in Africa countries?- is it every five years?

Regards.
