Subject: Merging data files
Posted by Elsha on Tue, 17 Sep 2024 21:56:51 GMT
View Forum Message <> Reply to Message

Hi,

I am working on merging injury data with IR, MR, and PR files. Can you help me please.
I had merged injury file with PR 1:1 in STATA by generating a new variable. for instance for injury
data: gen ID_new = hhid + "_" + string(haiidx) + "_" + string(hv003)

I am using data for Nepal Demographic and Health Survey 2022. The DHS Program - Nepal:
DHS, 2022 - Final Report (English)

The files I am using are NPPR82DT : shows 74 self-harm cases while NPAI82DT shows 86
self-harm cases. Is it because 8 suicidal deaths and 4 were not in household. Hence information
regarding those were not collected in PR data files?

I want to merge NPPR82DT with NPIR82DT, NPMR82DT and NPHR82DT.

I want to link the socio-economic factors like religion, ethnicity, occupation with self-harm data. But
are these variables only in IR data files?

However, on merging the dataset on self-harm shows 72 counts instead of 86 (originally in injury
dataset). Why is that so?

Please help me!

Subject: Re: Merging data files
Posted by Bridgette-DHS on Mon, 23 Sep 2024 11:47:39 GMT
View Forum Message <> Reply to Message

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

I have prepared a Stata program that constructs the AI file (attached here). The program also
merges it with the PR, IR, and MR files. This file may be a little different from the AI file you are
using, so I am suggesting a slightly different filename (without "FL"). You will want to save the
program as a do file and change the paths. I constructed my own AI file just so I would understand
it better.

The AI file includes some people who have died, and they will not be in the other files. It includes
some people who survived but are no longer in the household, and they too will not be in the other
files. It allows for more than one accident per person; there are 16 people with 2
accidents/injuries. There are females who are outside the age range for the IR file and for males
who are outside the age range for the MR file. I expect that some ages in the AI file do not match
exactly with ages in the other files. Those people who had 2 accidents or injuries are duplicated in
the merge.

An analysis of risk factors must include cases in the PR, IR, and MR files who did NOT have an accident or injury in the reference period. These case cases that did not match with the people in the AI file. I suggest constructing some recodes--for example a yes/no binary variable for "had an accident" and a yes/no binary variable for "had an injury".

In the PR merge, household-level data can be attached to people who died or are no longer in the household; I did not do that.

I am not directly answering your question about the self-harm cases but there are several potential issues with the analysis of the AI file and they go beyond the scope of the users forum.

File Attachments
1) accidents_injuries_do_20Sep2024.txt, downloaded 128 times