Subject: Inquiry regarding DHS 2018 analysis using R
Posted by woojae1995 on Wed, 18 Jan 2023 11:27:46 GMT
View Forum Message <> Reply to Message

I am currently doing a secondary analysis project using the 2018 DHS dataset of Nigeria.

Currently, I am using R and I have several technical/coding questions. I am currently using the 2018 DHS individual & children dataset.

1) How do I get a list of the column labels in R?
- I want to know the labels for the column (ex. b19 = current age of child in months) and the labels for the answer choices (ex. for the question asking the sex of the respondent; b4, 1= male, 2=female)
- I did find the 'STANDARD RECODE MANUAL for DHS-7' published by the USAID, but it still does not have the full response labels.
- Is this a problem inherent to using R? I heard that labels are easily visible when using STATA. However, since I have been using R till now, I wonder if there is a way to create a list of all the questions & labels for the dataset I am using.

2) How do I merge two dataset in R?
Referencing from this site 'https://dhsprogram.com/data/Merging-datasets.cfm', I merged the children dataset & individual dataset using the following code in R

NigeriaIR <- read_dta('NGIR7BFL.DTA')
NigeriaChildrenKR <- read_dta('NGKR7BFL.DTA')
NigeriaKRIR <- merge(NigeriaChildrenKR, NigeriaIR, by = c('v001','v002'))
*IR = individual dataset, KR = children dataset

Is this the correct way to merge it? I am concerned because the children dataset itself has 33924 observations, individual dataset has 41821 observations but when I merge them by v001 (cluster number) v002 (household number), I get 52982 observations.
From my crude understanding, I cannot understand how the merged dataset has more observations than the number of observations for the individual dataset. Could anyone explain why this is happening or what I am doing wrong?

3) Is this the correct way to account for the weighted-survey?
NigeriaKRIRsvy <- svydesign(id = NigeriaKRIR$v021.x, strata=NigeriaKRIR$v022.x, weights = NigeriaKRIR$v005.x/1000000, data=NigeriaKRIR)
*NigeriaKRIR is the merged dataset name
*for some reason, after I merged the dataset, the vXXX variables (ex. v001, v002) change to vXXX.x (ex. v001.x, v002.x)

Thank you all in advance