Subject: Mismatch When Merging KR and PR recode files due to disparities in age variables
Posted by RobertB on Tue, 18 Jun 2024 15:19:38 GMT
View Forum Message <> Reply to Message

Hello all,

I have extracted and merged variables of interest from the PR and KR recode files of Kenya's DHS 2022. However, i have noticed that some records in the
KR file are not in the PR file after accounting for scenarios elaborated in this post
https://userforum.dhsprogram.com/index.php?t=msg&th=1900 &goto=28815&#msg_28815.
Upon investigating the records, I notice disparities between the hv105 (household member's age in years) and hc1 (age of children in months) in the PR recode and b8(age of children under 5 in years) and b19(age in children under5 in months) in the KR recode.

For example, I found an instances where a children is recorded to be 5 yrs under hv105 but the corresponding age in months as per hc1 is 50 months indicating that the child should be classified as 4yrs under hv105. Consequently, I ended up with records in KR not matching with PR because I had subset the data in each recode files independently before merging. However, the hc1 value in PR reflects correctly in the b19 and b8 records in the KR record.

My question is which of these two variable/files are correct?
Secondly should i merge the two files first then subset children under 5 using b8/b19 in the KR assuming these are correct?
or should i update the hv105 based on hc1 as shown in the code below and proceed with the merging?

gen age_hsehold = hv105

replace age_hsehold=0 if hc1<12
replace age_hsehold=1 if hc1>=12 & hc1<24
replace age_hsehold=2 if hc1>=24 & hc1<36
replace age_hsehold=3 if hc1>=36 & hc1<48
replace age_hsehold=4 if hc1>=48 & hc1<60

Thanks,

Regards,
Robert B.

Subject: Re: Mismatch When Merging KR and PR recode files due to disparities in age variables
Posted by Bridgette-DHS on Thu, 20 Jun 2024 14:36:16 GMT
View Forum Message <> Reply to Message

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

Matching of records in different files should be done solely with ID codes. In the case of children, you match hv001 hv002 hvidx in the PR file with v001 v002 b16 in the KR file. There is no other reliable way to do a merge.

To get into the KR or BR files, a child must appear in the birth history of a woman in the IR file. Children who have died or are not living with the mother will appear in the KR/BR file but not in the PR file.

Children in the household whose mother is not in the household, because she has died or lives elsewhere, will appear in the PR file but not in the KR/BR file.

The b variables are forced to be internally consistent during data processing. For example, in the older surveys b8 is calculated from cmc of birth and cmc of interview as "b8=int((v008-b3)/12)". In the newer surveys b8 takes day of interview and day of birth into account. To show in two steps, "age_in_days=mdy(v006,v016,v007)-mdy(b1,b17,b2)" and b8=int(age_in_days/365.25)". (I am using the Stata date function mdy here.) The reported age of the child in the household survey, hv105, is ignored. It is not at all unusual for b8 and hv105 differ, and when they differ, priority is given to b8. Note that the information in the birth histories is provided by the mother. The information in the household survey is provided by the household informant (whose line number is given by hv003). The household informant may be someone other than the mother and be less informed than the mother. Even if the household informant is the mother, her responses in the birth history are considered to be more reliable than her responses in the household interview.

Hope this is helpful.

Subject: Re: Mismatch When Merging KR and PR recode files due to disparities in age variables
Posted by RobertB on Thu, 27 Jun 2024 16:56:23 GMT
View Forum Message <> Reply to Message

Yes this is very helpful and have successfully merged my data. Thanks for the detailed explanation.

My next step is to conduct analysis and I am using data from KDHS 2003, KDHS 2008_9, KDHS 2014 and KDHS 2022. I am aware that i need to account for the weights-complex survey design and have reviewed several posts in the forum around this topic. It seems there were mixed views around the need to weight data in regression analysis from the earlier posts but now more people are adopting the weights based on the questions posted. I also see some published work that have not accounted for weights. In my case I intend to fit a multilevel logistic regression at individual level.

Could you please confirm if the current recommended practise is to include the weights/survey

design? Additionally is there a DHS documentation that elaborates on how to apply the weights if i combine data from the 4 surveys?- from the forums it is not clear to me when to rescale the weights if pooling data across multiple surveys from one country.

Thank you.

Best wishes,
Robert B.

Subject: Re: Mismatch When Merging KR and PR recode files due to disparities in age variables
Posted by Bridgette-DHS on Fri, 28 Jun 2024 12:24:21 GMT
View Forum Message <> Reply to Message

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

DHS strongly recommends using weights, as well as the other svy adjustments for clusters and strata. If you include weights, the estimates become unbiased. The other adjustments provide robust standard errors because they take into account the stratified two-stage sample design.

When you pool the four surveys into a single large file, you need to construct a variable to distinguish the surveys, for example survey=1, 2, 3, 4. The clusters and strata have to be re-numbered to distinguish all four surveys. For example, you could have "egen clustered=group(survey v001)" and something similar for strata. (The stratum code is v023 in the two most recent Kenya surveys but may be different in the earlier two.) The svyset statement would be something like "svyset clusterid [pweight=v005], strata(stratumid) singleunit(centered)". Note that in this statement v005 is NOT altered. This will be ok for all analyses I can think of EXCEPT for analyses that do not distinguish the surveys. For example, staying with the original weights would be questionable if you tried to calculate the mean of v201 (children ever born) in the four surveys, because that mean would be biased toward the largest of the four samples, and that's not desirable. But I would say that there is no reason to calculate such a mean. The mean of v201 (for example) should be calculated within each survey, but not for all surveys pooled, because you need a reference date for each estimate of v201.

You can find more discussion of weights in the Guide to DHS Statistics (https://www.dhsprogram.com/Data/Guide-to-DHS-Statistics/inde x.cfm), in the FAQ for the user forum, and within the forum itself, if you search topics and keywords.

Subject: Re: Mismatch When Merging KR and PR recode files due to disparities in age variables
Posted by RobertB on Mon, 08 Jul 2024 13:42:02 GMT
View Forum Message <> Reply to Message

This is very helpful. Thank you for your feedback.

Best regards,
Robert B.