Posted by Jeremy on Thu, 31 Aug 2023 09:31:16 GMT
View Forum Message <> Reply to Message

Dear DHS specialists,
    I have a question; I am doing research on Acute Respiratory Infection using the EDHS dataset. I merged the KR file with PR file. For the analysis I use the mixed effect multilevel regression model and do the analysis using R. I adapted the complex sample design provided by your office "DHSdesign<-svydesign(id=dta$v021, strata=dta$v023, weights=dta$wt, data=dta)", as follows:
ARI_TRA1 <- svydesign(id=~V021, weights=~WGT,strata=~V023, nest=TRUE, survey.lonely.psu = "adjust", data=ARI1st)
However, the sum of the estimates in the weighted frequency calculation (9904.89) is not the same as the original sample size (8781). (Descriptive analysis of the alive children aged below 60 months). Why is this difference, please help. Stay Blessed! Stay Safe!
Unweighted frequency
ARI4th %>% freq_table(V024)
Result:-
  Freq. %
0 1650.97 16.67
1 2129.04 21.49
2 2002.36 20.22
3 2168.62 21.89
4 1953.90 19.73
n_total 9904.89 100

weighted frequency
outp1 <- svytable(~Child_age,design=ARI_TRA1)
outp2 <- round(prop.table(svytable(~Child_age,design=ARI_TRA1))*100,d igits=2)
cbind(outp1,outp2)
Result:-
  var cat n percent
1 Child_age 0 1443 16.43
2 Child_age 1 1821 20.74
3 Child_age 2 1815 20.67
4 Child_age 3 1923 21.90
5 Child_age 4 1779 20.26
n_total   8781 100.00

Subject: Re: Need help
Posted by Bridgette-DHS on Thu, 31 Aug 2023 13:15:10 GMT
View Forum Message <> Reply to Message

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

You do not have a problem. First, you seem to expect that the weighted and unweighted numbers of cases should be equal--or equivalently, the mean weight is 1--but that's not necessarily the

case. I suggest that you open the KR file and enter these two Stata lines (or their equivalent in R):

summarize v005
summarize v005 if b16>0 & b16<.

You will see that the mean weight is not 1 (or 1000000) for the KR file as a whole or for the subset of children who are in the PR file. That's because v005 is normalized during data processing ONLY to have a mean of 1 (1000000) in the entire IR file. Then the mother's weights are assigned to the children. Women can have any number of children. Subgroups of children, or of women, will therefore have mean weights that differ from 1.

Second, when you estimate a statistical model, at least in Stata, using pweights, Stata will automatically re-normalize the weights so that they have a mean of 1. That's a default that makes sense and that I don't think you could override, even if you wanted to. I expect that R does the same thing.

So, as I said, you do not have a problem.

Subject: Re: Need help
Posted by Jeremy on Thu, 31 Aug 2023 14:44:57 GMT
View Forum Message <> Reply to Message

Thank you so much Sir and Dear Madam Bridgette. Have a Great Day. Stay Safe!

Blessings!

Jeremy