Subject: Balancing number of observations for the dependent and independent variables
Posted by I am Faithful on Tue, 07 Mar 2023 02:54:32 GMT
View Forum Message <> Reply to Message

Hello

I am doing a research on maternal empowerment and the effect on underfive child nutrition in Malawi, I am using KR dataset.
With that in mind, I am trying to do some inferential and descriptive statistics to identify significant variables to use in the final model but for some variables, the number of observations do not balance.

For instance, when I do

Please assist how I can balance the two

Subject: Re: Balancing number of observations for the dependent and independent variables
Posted by Janet-DHS on Tue, 07 Mar 2023 20:25:07 GMT
View Forum Message <> Reply to Message

Could you provide the survey year you are referring to?

Subject: Re: Balancing number of observations for the dependent and independent variables
Posted by I am Faithful on Wed, 08 Mar 2023 00:39:10 GMT
View Forum Message <> Reply to Message

Thank you Janet.

I am using DHS Malawi 2015/16.

I have a number of independent variables are residence, maternal education, child age, child sex, sex of household head, maternal occupation, among others. I have also created a maternal autonomy and maternal knowledge acquisition index as as main independent variables of interest.

The dependent variables are stunting, wasting and underweight.

So I want the number of observations in these variables to balance.

Subject: Re: Balancing number of observations for the dependent and independent variables
Posted by I am Faithful on Wed, 08 Mar 2023 15:41:11 GMT
View Forum Message <> Reply to Message

Hi Janet

Let me clarify my point to say I meant the

summarize command shows different number of observations between for instance stunting and child gender, age of household head and not Tab command as earlier stated.

Again, is it ok to proceed with the analysis if summarize command shows different number of obs between the dependent and independent variables?

Subject: Re: Balancing number of observations for the dependent and independent variables
Posted by Janet-DHS on Wed, 08 Mar 2023 21:47:58 GMT
View Forum Message <> Reply to Message

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

I believe you are doing several different regressions, and are getting different sample sizes (n's). This can happen because the different variables may have different numbers of cases that are not applicable or are automatically excluded for different reasons. If you want all the models to have the same number of cases, then you have to define a variable "varsmissing" (for example) that is coded "1" if a case is dropped from ANY of the models and "0" otherwise. Then you re-run the models with a line "if varsmissing==0". There are alternative ways to do this, for example with "svy: subpop(X)". (If you do it with subpop, the variable X in parentheses should be 1 if you want to KEEP the case, the reverse of the coding I suggested for "varsmissing".)

There are advantages to having the same n for several models, for example if you want to test one model against another. But if you lose a lot of cases from just one or two of your variables, it may be preferable to drop the variable and keep the cases.