
Subject: Using weights for merged PR&IR file NFHS-4(2015-16), India

Posted by [preshit](#) on Mon, 11 Feb 2019 23:00:14 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello All,

I am doing decomposition analysis using a logit model for which I have merged NFHS-4(2015-16) PR and IR recode files. I am using STATA module for Fairlie decomposition technique. My main dependent variable is RSBY enrollment taken from PR file while the main predictor variable is caste status for the household head which is also taken from PR file. My other predictor variables are taken from both PR and IR files. The predictor variables I included from IR file are- women's and husband's occupation and chronic disease status which are designed to provide information at state level only as these questions were asked to subsample of women in state module (From NFHS-4 report pg.1: NFHS-4 was designed to provide information on sexual behaviour; husband's background and women's work; HIV/AIDS knowledge, attitudes, and behaviour; and domestic violence only at the state level (in the state module)).

My analysis is confined only to observations which are present in the IR file i.e. eligible women. I wish to conduct my decomposition analysis using sample weights now. I have gone through previous forum postings and web-pages and YouTube videos posted; however, I am still not clear which weights to use in my case. The closest answer I could get is posted in the following link (screen-shot is attached as well)-<https://userforum.dhsprogram.com/index.php?t=msg&goto=16014&S=Google>

As suggested by the Statisticians at the DHS program in the above link, I used weights from IR file(v005) as my model included few variables from IR dataset. However, I am getting weird results using these weights and the group difference explained by variables included in the model is going above 100% which doesn't make any sense. My question is that the pweight created using v005 is still correct weight in my case or I should be using some different weight for analysis? e.g. hv005?

Thank you so much in advance for your help.

File Attachments

1) [India weights SAS and merging.jpg](#), downloaded 701 times

Subject: Re: Using weights for merged PR&IR file NFHS-4(2015-16), India

Posted by [Bridgette-DHS](#) on Tue, 12 Feb 2019 16:01:30 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I just merged the IR and PR files for the survey. There are 699,686 women in the IR file and all of them are in the PR file. The correlation between v005 and hv005 is 0.9982. Switching from v005 to hv005 for your weights cannot affect the problem you are having. I doubt that the problem has anything to do with the weights. Have you tried the decomposition without using weights at all? That's something to check.

In the PR file, the RSBY variable is sh55d. It is binary (0/1). I have not used the Fairlie decomposition but I see that it is specifically for binary variables. The caste variables, sh35 and sh36, respectively, have 4 and 5 categories, respectively. There are $4 \times 5 = 20$ combinations. There are 20×2 combinations with sh55d, and I see that none of them are empty, so your problem is not due to empty cells.

Is the Fairlie method designed for categorical predictors? Could that be the problem?

My recommendation, whenever you have an analytical problem, is that you simplify the data setup as much as possible. But a bigger question is whether you need to apply the Fairlie method, or any other decomposition method, in this setup. You can easily identify which categories or combinations of sh35 and sh36 have high enrollment in RSBY and which ones have low enrollment, just using cross-tabulations. If you want to account for other variables, you can use logit regression. I wonder what other forum users would suggest.

Subject: Re: Using weights for merged PR&IR file NFHS-4(2015-16), India
Posted by [preshit](#) on Thu, 14 Feb 2019 18:26:17 GMT

[View Forum Message](#) <> [Reply to Message](#)

Tom,

Thank you so much for your reply. My primary question was on which weights to use in my case and you have clarified my doubts. Regarding my logit models, I have created caste dummies on sh35 and running separate models for each of them. You are right, Fairlie decomposition is for binary dependent variable and that's why I have created caste dummies. When I rechecked my codes, I found that my problem was due to estimation command ado file and sorting I did while creating new variables. I was able to resolve it now. Thank you so much once again for your prompt reply. I really appreciate it.

Regards
Preshit