
Subject: Response rate and weights

Posted by [shreyaj7](#) on Thu, 30 Mar 2023 07:30:13 GMT

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

It is mentioned in "Demonstration of How to Weight DHS Data in Stata" that when using couples' information one should use Men's weight as they have higher nonresponse rates. I am studying the effect on child health and want to use the sample in which the information for both the parents (like education, age, employment, etc.) are available (which is approximately 15% of the total sample in the KR file).

My concerns are:

1. Can we use this sample and still get results that are representative?
2. The weights that need to be used is men's weights or women's weight as mentioned earlier they have low response rates than women's.

Kindly help me out.

Subject: Re: Response rate and weights

Posted by [Bridgette-DHS](#) on Thu, 30 Mar 2023 15:06:42 GMT

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

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

The first issue I see is that children are linked with their mothers, and you cannot be sure that the mother's current partner is the child's father. To confirm that, you have to work with hv113-hv114 in the PR file, which identifies the father if he is alive and in the same household. Second, often there is subsampling of men for the interview with men. For example, in the NFHS-4 and -5, only 1/6 of men were interviewed. For these reasons, a study that includes the effect of the father's characteristics on the child's health and welfare can be challenging. (Important, but challenging.)

It is recommended that if you have a table, regression, etc., that includes variables from the survey of men, even if it also includes variables from the survey of women, you should use mv005 for the weight. The reason is that nonresponse is higher for men than for women. This would apply even if you are not using the CR file. It applies if you do any merge with the MR file and are including in your command any variables from the MR file.

At the same time it should be said that the estimates will not be very sensitive to which weight you use. Your conclusions will probably be robust with respect to the choice of weight. It's just considered to be "best practice."

Subject: Re: Response rate and weights

Posted by [shreyaj7](#) on Fri, 31 Mar 2023 04:14:30 GMT

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

Thank you so much for your reply.

Just to make sure I understood it correctly. If I merge the data from KR, MR and IR files and exactly get the information on the parents of the children and do my analysis. Though the sample size would drastically reduce to almost 1/6 of the total sample I can generate tables and do regression by applying mv005 weights. Can I still get results that will be representative? or it doesn't matter for causal inference for it to be representative?

Any help is appreciated. Thank you.

Subject: Re: Response rate and weights

Posted by [Bridgette-DHS](#) on Fri, 31 Mar 2023 16:01:13 GMT

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

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

Recently I posted a program to merge all the data for children, mothers, and fathers. It may go beyond what you need, but it combines the data in the BR ((and KR), PR, IR, and MR files with children age 0-17 as cases. For the NFHS-4 and -5 surveys, which had only a 1/6 subsample of men, you will lose a lot of children and mothers in order to get the fathers, but yes, you will still have a representative sample. The estimates will be unbiased. And because these surveys were so large, you will still have a large sample.

If you want to compare the significance of effects for the mothers and fathers, you need to be careful. To take a simple example, say you wanted to look at the effects of maternal and paternal education on child survival. If you use the full sample to estimate the maternal effect, and the 1/6 sample to estimate the paternal effect, both coefficients will be unbiased, and comparable. However, even if the effects were equal the t or z score for the mothers would be about $\sqrt{6}=2.4$ times as large as the one for men, with much more potential to be statistically significant. You'd have to take that difference in statistical power into account if you inferred that the mother's education was significant, but the father's education was NOT.

File Attachments

1) [merge_children_mothers_fathers_do_14Feb2023.txt](#), downloaded 120 times

Subject: Re: Response rate and weights

Posted by [shreyaj7](#) on Sat, 01 Apr 2023 08:52:33 GMT

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

Thank you for sharing the do file. It's very helpful. If at all possible for you can you share the

merge_children_mothers_fathers.dta file. I want to compare my merged data file with it if I have got it right or made some mistakes in the process.

Please clarify one thing for me. In your example "effect of maternal and paternal education on child survival" you are looking at the effects of mother and father in separate regressions, right? Not taking them together in one regression? because if we do then we will be left with near about those children observations only for whom we have both parent's info who are alive and live with them. so roughly 35-40k observations.

Also, I had one more query. I just need father's education and mother's education to create my variable of interest and other controls could be mother, child, HH characteristics. So can I use the info about the child's father line no. in PR file and from there merge into KR file?

I tried this as well and now I have father's education variable for approximately 1,73, 000 children (in KR file out of 2,32,920).

Can I now use these 1,73,000 observations as my sample size and apply women's weights and do my analysis? Will that be correct?

Subject: Re: Response rate and weights
Posted by [Bridgette-DHS](#) on Mon, 03 Apr 2023 13:15:16 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Yes, you can get father's education from hv106 on the father's line in the PR file, and you don't need mv106 in the MR file. That's good because there are about 6 times as many men in the PR file as in the MR file (in NFHS-5). hv106 is considered to be less accurate than mv106, because the information comes from the household respondent, who may be someone other than the man himself, but it's useful for your purposes.

In your merged file I would agree that v005 would be the weight to use--it's the weight for the mother and also for the child--except that if you go ahead and merge with the MR file as well, you should use mv005 when your model includes variables from the MR file. That will give the best adjustment for nonresponse.

I don't have time to construct that merged file, and in any case it would be too big to send. The program is really just a template that users can adapt. You may want to do some checks to make sure it contains the cases and variables you want.

Subject: Re: Response rate and weights
Posted by [shreyaj7](#) on Mon, 03 Apr 2023 17:41:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much for your response. I used your commands and have constructed the data for children and parents in one file. I have now $n=920461$ (all ages) and 3026 variables. Hope that's correct. I am doing the checks as you mentioned to make sure it was done right.

I just need to confirm this:

As I need data for children aged between 0-5 years who live with both of their parents and are single births. The sample size would be around 1,70,000. So, when I will use this analytical sample further in my analysis, will the weights be still representative or it wouldn't matter if we use or don't use weights as they are not any more representative at the national or subnational level?

I am sorry for coming back to representativeness question again and again.

Your response is always very helpful. Looking forward to another one. Thank you

Subject: Re: Response rate and weights
Posted by [Bridgette-DHS](#) on Mon, 03 Apr 2023 19:01:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You still need to use weights in all analyses in order to have unbiased estimates of coefficients. Weights compensate for the over-sampling of some sub-populations and the under-sampling of some other sub-populations. If you do not use weights, all statistics will be biased toward the over-sampled sub-populations and away from the under-sampled sub-populations. The merging has not altered the need to use weights.

As I have said elsewhere, some people, mainly economists, do not like weights, and they can be very dogmatic about that position. I am not being dogmatic. I just put a high value on having unbiased estimates.

Subject: Re: Response rate and weights
Posted by [shreyaj7](#) on Mon, 03 Apr 2023 21:01:24 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much for your patience and all your help Tom.

Subject: Re: Response rate and weights
Posted by [shreyaj7](#) on Wed, 12 Apr 2023 08:08:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Tom & Bridgette,

I am making a table for the prevalence of hypogamous marriages (Women marrying men with lesser education than them) state-wise for India.

The following steps were taken:

1. I used IR women's recode> selected only those women whose husband's background characteristics are available(1/6th sample N=81487)> created variable hypogamy [Binary variable takes value 1 if wive_edu> Husb_edu; 0 otherwise]
2. set svy> svyset v021 [pw=wgt], strata(v023) singleunit(centered) ,where wgt= v005/1000000
3. Now I used the syntax to get my proportions> svy: mean hypogamy, over(state) , this gives me the percentage of hypogamous marriages in each state and all India average as(25.98%)
4. I repeated the exercise by using state weights(sweight) and by merging men's weight from men's recode (mv005). I got the same results for state level percentages by using state weights but different All India average (24.91 %). Similarly using men's weight changes the percentages for states and as well the all India average (27.26%).

Now I want to know if got the steps right and which results to go with I mean which weights to use in this context. Looking forward to reply. Thank you.

Subject: Re: Response rate and weights

Posted by [Bridgette-DHS](#) on Wed, 12 Apr 2023 11:48:29 GMT

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

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

This looks good to me. The state weights are equal to the national weights multiplied by a constant (for each state) so the means should come out the same using state weights or the national weight. The national weight takes into account the different sampling fractions in different states and is definitely what you want to use for national estimates.

The only thing I might do differently would be to use the CR file rather than the IR file. For couples in the MR file, the woman and man have to name each other as partners, leading to better matching. The level of education for the man is reported by the man himself, rather than from the wife, who may introduce some bias, especially if there is a large difference in their levels of education. If you repeat the analysis using the CR file, the man's weight would be preferable because there is a higher level of nonresponse for men.

I would also include parallel analyses of hypergamy (marrying up) and homogamy (the same level). I think what you want to get at is the balance between hypogamy and hypergamy, and that's going to be affected by the amount of detail in the education distribution. For example, if the distribution is very coarse, such as no/any education, then that alone will lead to more homogamy and less of the other two.

Subject: Re: Response rate and weights
Posted by [shreyaj7](#) on Thu, 20 Apr 2023 00:50:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Tom so much for your response.

I am unable to locate a variable that tells whether the wife and husband are of the same caste or not. Do we not have that variable in NFHS-5 (India)?

Please let me know. If we do have which is the variable measuring this.

Thank you.

Subject: Re: Response rate and weights
Posted by fred.arnold@icf.com on Thu, 20 Apr 2023 15:21:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

Women and men who are eligible for the individual questionnaire are each asked what their caste/tribe is and whether they belong to a scheduled caste, a scheduled tribe, an other backward class, or none of these. However, although the first question specifies the caste/tribe, there are more than 1,000 castes/tribes and there is no variable for those castes/tribes. Also, a much smaller percentage of men than women are eligible for an individual interview.
