
Subject: Interpretation of Rescaled household level weights for India-NFHS4

Posted by [preshit](#) on Mon, 20 Jan 2020 16:28:21 GMT

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

Hello DHS Forum Members,

My apologies in advance for lengthy post but I want to explain the issue in detail so it will be useful to others as well. For my analysis, I am using the NFHS-4 PR file and multinomial logit regression. I am stratifying my analysis on rural and urban samples and using svy command for population-level inference. I have rescaled my weight variable and my Stata code looks like below:

```
gen newhv005= hv005/1000000
```

```
svyset,clear
```

```
svyset [pw=newhv005],psu(hv021)strata(hv023) singleunit(centered)
```

```
svy, subpop(if respondent_residence==0) : mlogit Y X1##(i.X2 i.X3) i.X4 i.X5 i.State Fixed Effect  
// for rural sample  
mlogit,rrr
```

```
svy, subpop(if respondent_residence==1) : mlogit Y X1##(i.X2 i.X3) i.X4 i.X5 i.State Fixed Effect  
// for urban sample  
mlogit,rrr
```

```
svy : mlogit Y X1##(i.X2 i.X3) i.X4 i.X5 i.respondent_residence i.State Fixed Effect // for entire  
sample  
mlogit,rrr
```

I have tested my code with and without svy setting my data. Without svyset all mlogit regressions give me coefficients, SEs, CIs, and associated P-values. However, when I svyset my data, for urban sample, I am getting the following error:

Warning: variance matrix is nonsymmetric or highly singular

And the output is missing SEs, CIs, and associated P-values.

I searched the Stata user forum and realized this error is because one of my Strata has only one PSU which can be checked with the command:

```
svydes if respondent_residence==1 //to check if there is only one PSU within any strata
```

I approached Stata Technical Support and received the following reply:

I have checked your data, the pweight you are using, newhv005, has a relatively big variation. The minimum value is .000673 and the maximum is 38.9548. If the computation has perfect precision, there will not be a problem. However, all the computation on computers have limited

precision, and -mlogit- is particularly sensitive to variation in pweight. The missing value you saw is a numerical problem because of limited precision. You should check your data to see why the pweight has such a big variation.

If you still want to use the same pweight variable, you can avoid this numerical problem by rescaling your pweight. Instead of dividing hv005 by 1000000, you can divide it by 100000000 or larger number. This will solve the problem of missing SE. Please note that rescaling the pweight does not affect the point estimates or standard errors of the coefficients. It only affects the population size reported. In this case, you should use a different unit to interpret the population size reported, e.g., instead of using one person, you should use a group of 100 persons as the unit.

Given this reply I have the following questions:

1. How the interpretation of -mlogit- will change if I rescale pweight to 10000000 and not to the recommended 1million?
2. Or shall I rescale the pweight in some other way?
3. Is there any other way I should specify my model?

Thank you in advance for your time and for reading the post.

Regards
Preshit

Subject: Re: Interpretation of Rescaled household level weights for India-NFHS4
Posted by [Bridgette-DHS](#) on Fri, 24 Jan 2020 15:00:16 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The huge variation in national weights for the NFHS4 is unfortunate, but it arises mainly because of the huge variation in the populations of the states and districts. Below I will paste a list of the mean weight (hv005/1000000) in each state, ordered from smallest to largest. Within each state, there is considerably less variation in the weights.

I agree with your specification of svyset and the use of subpop() with svy. The advice from Stata Technical Support is good but I really don't think re-scaling the weights will eliminate the problem, because Stata automatically re-scales pweights so they add to 1. The location of the decimal point in the weight is irrelevant. You can define wt=hv005 or wt=hv005/1000000 or wt=hv005/100000000 or use any other multiplier or divisor for hv005 and the results (with pweight) will not change. I suggest that you try that, but as I said I doubt that it will eliminate your problem.

I suggest that you add this line right after you load the data: "recast double hv005". That will

change hv005 to double precision, and I believe all calculations involving hv005 will then include many more decimal places, in effect averting a matrix calculation which involves division by zero, which is what causes singularity in the variance/covariance matrix. The only alternative I can suggest is to drop the cases with the very lowest values of hv005, because their impact on any national results will be negligible, and that's another way to avoid dividing by zero. Please try this and let us know whether it works.

File Attachments

1) [HV024.JPG](#), downloaded 976 times

Subject: Re: Interpretation of Rescaled household level weights for India-NFHS4
Posted by [Bridgette-DHS](#) on Tue, 28 Jan 2020 21:01:49 GMT

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

Are you still getting the Stata data error message?

Subject: Re: Interpretation of Rescaled household level weights for India-NFHS4
Posted by [preshit](#) on Tue, 28 Jan 2020 21:13:32 GMT

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

Ms. Bridgette and Dr. Pullum,

Thank you so much for your useful suggestions. My apologies for the delayed response. I tried all the suggestions given by Stata Tech Support and you but issue did not resolve. I tried excluding some states as well having lowest values of the weight variable but issue persists with the urban subsample. So, I have changed my model specification as follows:

```
svy : mlogit Y X1##(i.X2 i.X3 i.respondent_residence)) i.X4 i.X5 i.State Fixed Effect  
mlogit,rrr
```

then I interpret all three interaction terms for all my outcomes. The weight variable I use is wt=hv005/1000000. I was reading somewhere that interaction model is more useful than stratified models, so I am OK with this approach. I am getting the output and standard errors are not inflated. Please let me know if you have found solution for urban/rural stratified analysis.

Regards
Preshit

Subject: Re: Interpretation of Rescaled household level weights for India-NFHS4

Posted by [Bridgette-DHS](#) on Tue, 28 Jan 2020 23:35:45 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I'm glad you have worked out a solution. By coincidence someone here at DHS had a similar problem recently, with a different survey, for which the weights had much less variation, so it was clear that the weights are not the problem. Most likely the problem is due to combinations of variables that are empty, that is, have no cases. I believe that your specification of the interactions has bypassed the problem and could help other users.

Subject: Re: Interpretation of Rescaled household level weights for India-NFHS4
Posted by [preshit](#) on Tue, 28 Jan 2020 23:53:03 GMT

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

Yes, initially I was suspecting this possibility as well. By dropping and including various combinations of variables, I realized that my interaction terms were creating such combinations of empty cells. However, I wanted to retain the interaction terms and was looking for alternatives. Stata Tech Support directed me to the weight precision issue. I have suggested a possible alternate model specification in my previous post. If you deem feasible please share the approach your team person adopted to resolve the issue. By doing so, the forum users will have both the alternatives available if anyone encounters similar issue. Thank you.
