
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
