Subject: domestic violence weights at psu level
Posted by akarshik on Fri, 17 Mar 2023 21:39:02 GMT

I am running a selection on observables model using the DHS datasets of India.

In my model, I calculate a covariate measure at the primary sampling unit (PSU)level. My binary dependent variable is whether intimate partner violence occurred in the past 12 months. I understand that the domestic violence module weight makes the model results nationally representative. But since my covariate measure is at the PSU level, I calculated an adjusted weight per household by taking a ratio of the national domestic violence of that household to the average of the national domestic violence weight at the PSU level.

Can you please let me know if this approach is okay? If not can you please suggest an alternative approach?

Subject: Re: domestic violence weights at psu level
Posted by Bridgette-DHS on Mon, 20 Mar 2023 13:47:04 GMT

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

I'm not clear how you are calculating your PSU-level binary variable. Is it 1 if ANY woman in the PSU (cluster) reported IPV in the past 12 months, and 0 otherwise? Why would you define the covariate this way? If a cluster has more women in it, the probability that one of them will report IPV is greater.

If the cluster is the unit of analysis, then you need the cluster-level weight. I agree that it is not hv005, but it is complicated to calculate. We have had several postings on multi-level weights, including for the India surveys. These describe how to separate hv005 into a person-level weight and a cluster-level weight, the product of which is hv005. It sounds like you only need the cluster-level weight, but it's not easy to get.

With DHS data, the cases are individuals--household members or women or men. You can also use households as units. I recommend that you try to formulate your model so individuals (or households), rather than clusters, are the units. Otherwise you are not making full use of the data. Can you provide more explanation of your approach?

Subject: Re: domestic violence weights at psu level
Posted by akarshik on Mon, 20 Mar 2023 14:12:19 GMT

Thank you so much for your reply.

My dependent variable is a binary measure of IPV. It is at the household level.

All my covariates are also at the household level except one. Let us call that covariate X. Covariate X is calculated as follows:
I take the PSU level average of column A. Then at the household level, I take a ratio of column A to the PSU level average of column A. For example - if column A is the number of children in a household, then covariate X will tell me if a household has more than the average number of children per household in the respective PSU.

So I was confused if I should use hv005 or d005. I feel like I should convert the weights to a cluster level. Please let me know your suggestions on this. I am very grateful for your help.

Subject: Re: domestic violence weights at psu level
Posted by Bridgette-DHS on Mon, 20 Mar 2023 15:35:28 GMT
View Forum Message <> Reply to Message

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

Thanks for the clarification. You do not need to alter the weight. Just use d005, as it is. The inclusion of the cluster-level covariate, as you have described it, attached to the individuals or households, does not require a multi-level model or any change to the weights.

Subject: Re: domestic violence weights at psu level
Posted by akarshik on Mon, 20 Mar 2023 16:12:55 GMT
View Forum Message <> Reply to Message

Thank you so much for clarifying.

I have a follow up question.
Is it important to use d005 at all? I am using the cluster(psu) command in STATA, to cluster my standard errors at the PSU level. Does that suffice for the model's robustness?

Subject: Re: domestic violence weights at psu level
Posted by Bridgette-DHS on Mon, 20 Mar 2023 19:30:04 GMT
View Forum Message <> Reply to Message

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

We recommend using svyset and svy for your final analyses, at least. Your svyset command

would look like this: "svyset v001 [pweight=d005], strata(v023) singleunit(centered)". If any of your variables come from the DV module, d005 is preferable to v005, because it is adjusted for nonresponse to the module. The svy adjustments, and the use of d005 in place of v005, will not have a huge effect on your results, but they are strongly recommended.

Subject: Re: domestic violence weights at psu level
Posted by akarshik on Tue, 21 Mar 2023 04:01:16 GMT
View Forum Message <> Reply to Message

Thank you so much for your help.

In my linear regression with the dependent variable as IPV, I have one covariate X, as explained earlier, which is a ratio of column A to the psu level average of column A. I also have two more important co-variates B and C. I specify a model :

reg IPV B##C##X more controls // code in stata

The variables IPV, B, C, X are all binary 0,1 variables.
In the unweighted model, I get significant results for my two-way and three-way interaction terms with X and the two variables, say B and C.
However, the interaction terms (two-way and three-way) are no longer significant if I use the command svy with svyset v001 [pweight=d005], strata(v023) singleunit(centered).

I suspect that since psu-level comparisons are happening in variable X, using the national-level survey weight d005 may not be the best approach. Previous literature suggests that I should at least get significant two-way interaction between B and C, which I am not getting upon weighting in my case. I wonder if cluster-level weights would be more appropriate. Can you provide me the code to get cluster-level weights?

Can you please help me understand the best way to approach this issue? Thank you so much.

Subject: Re: domestic violence weights at psu level
Posted by Bridgette-DHS on Tue, 21 Mar 2023 11:39:21 GMT
View Forum Message <> Reply to Message

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

The svyset component to adjust for clustering tends to increase the standard errors--that is, to increase the width of confidence intervals and make test statistics less significant. However, that is a correction you need to make. The adjustment for weights corrects for bias due to the over- and under-sampling of subpopulations, mainly the different strata. The adjustment for strata has an effect that counter-acts the effect of the adjustment for clustering. All three things are determined by the two-stage stratified cluster design of the survey.

It can be disappointing when an effect or interaction in your model is not significant, but that happens all the time. There can be many reasons, including the cross-sectional nature of the surveys and the limited sample size for some analyses. Omitting the svy adjustments in order to get better results is not an option.

Subject: Re: domestic violence weights at psu level
Posted by akarshik on Tue, 21 Mar 2023 14:37:17 GMT
View Forum Message <> Reply to Message

Thank you very much for clarifying this.

Subject: Re: domestic violence weights at psu level
Posted by akarshik on Tue, 21 Mar 2023 19:09:04 GMT
View Forum Message <> Reply to Message

Hello,

I appreciate your help with the survey weights, and I have a follow-up question.
In case I want to use a propensity score matching method, do you recommend I use survey weights?

Suppose my dependent variable is IPV, treatment is treat, and I have a list of covariates, that can be used for matching and eventually linear regression. Can you please suggest the best code for the calculating the propensity scores?

I am confused if that step needs to have survey weights. Or if the scores get calculated and then it gets multiplied with d005 to give a weight I eventually use for linear regression. I greatly appreciate your help in this. Thank you.

Subject: Re: domestic violence weights at psu level
Posted by Bridgette-DHS on Wed, 22 Mar 2023 12:02:07 GMT
View Forum Message <> Reply to Message

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

Weights are not used in the construction of a variable--only in the analysis, as part of the estimation commands. If you use propensity scores, factor analysis, glm models, etc., you should be able to include a specification of weights with svyset and svy in the estimation command.

There are a few complex estimation procedures that do not have an option for weights. Historically, when a package just as Stata first includes a new method, it may not initially include

an option for weights, but in later versions the option is added. If there is no option for weights, then you have no choice, but if there is such an option, it's best to use it. Propensity scoring has been around for a long time and I'm pretty sure Stata allows pweights and svyset/svy for this procedure.