Subject: Multilevel model with region random effects
Posted by sophia614 on Thu, 16 Feb 2023 14:08:37 GMT

Dear DHS survey specialists,

I am using the women's data in Benin 2018 to run a multilevel model with region random effects in Stata. Most of my variables are at the individual level; however, I do have one measured at the regional level. I tried running the following commands in Stata, but received an error message after the second command (see below).

svyset psu [pw = weight], psu(psu) strata(v023) singleunit(centered)
svyset: xtmixed dep ind $controls || region_recent:

xtmixed is not supported by svy with vce(linearized); see help svy estimation for a list of Stata estimation commands that
are supported by svy

I believe that I need to create level weights, but am not sure how to do so. Do I create them at the region-level (since my random effect is at the region level) or at the psu level?

Any help you can provide is greatly appreciated.

Best wishes,
Sophia

Subject: Re: Multilevel model with region random effects
Posted by Bridgette-DHS on Fri, 17 Feb 2023 13:54:52 GMT

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

In the setup for a multilevel model, level 1 would be the household and level 2 would be the psu (cluster). In 2022 we prepared a methodological report describing how to do this (https://www.dhsprogram.com/pubs/pdf/MR27/MR27.pdf). There have been several postings on the user forum which will help with this subject.

However, if your higher-level variable is (or variables are) at the level of the region, I don't think you need a multilevel model. If you had cluster-level variables, yes, but region is not part of the sample design. An individual-level model should be sufficient.

Subject: Re: Multilevel model with region random effects

Posted by Aline Semaan on Wed, 29 Mar 2023 23:59:23 GMT

Hi Tom,

I am facing a similar situation as Sophia and I found your answer to be quite logical. But I have seen some contradicting information in a previous DHS analytical study (here: https://dhsprogram.com/pubs/pdf/AS26/AS26.pdf) and I would like to get a clarification.

I am conducting an analysis where I plan to study associations between health system capacity (extracted from SPA survey) and length of stay in health facilities after birth (from the DHS). The SPA survey for this country is not a census, hence only a sample domain linkage is possible (DHS Spatial analysis report 10). Thus I plan to link the two databases using the administrative level variable (i.e. Region), and consequently conduct a multi-level modeling approach since the outcome is at the individual level, but some predictors are at the regional level. To quote Wang et. al : "Using the standard statistical approach, questions arise about the appropriate unit of analysis. Analyzing data at the individual level ignores the nesting of people within regions, which results in underestimating the standard errors and increasing the chance of incorrectly rejecting null hypotheses (Raudenbush and Bryk 2002). Alternatively, if the unit of analysis is the region, it will be difficult to include individual-level variables in the analysis. We can address these problems with multilevel modeling and allow for simultaneous investigation of the effects of group-level and individual-level predictors on individual-level outcomes."

In their report, Wang and colleagues do not go into the details of adjusting for sampling weights, and you might be familiar with the fact that running an melogit in Stata fails when the level2 variable is Region because there are no weights for this level.

I failed to find solutions online up to now. Do you have any suggestions on how to conduct this analysis without ignoring the hierarchical structure of the data, and still correctly adjusting for sampling weights? Would including "region" in the multivariable model be sufficient in this case?

Many thanks,
Aline

Subject: Re: Multilevel model with region random effects
Posted by Bridgette-DHS on Fri, 31 Mar 2023 11:57:24 GMT

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

I have checked with Senior DHS Researcher, Sara Riese, and we agree that you could use a two-level model with women at level 1 and region at level 2, with 1 as the level 2 weight, because regions were not sampled. If the SPA you are using was not a census, and the facilities were sampled, then there were probably different sampling fractions for different facility types. The facility weights are proportional to the inverse of the sampling fractions for the respective facility

types. The client weight accounts for the selection of the facility.

Let us know if you have further questions. For your preliminary analysis, a simplified version of svyset will be ok, but you want to be confident in the final version.