Subject: Accounting for cluster variation
Posted by zaeema on Tue, 25 Aug 2015 17:53:11 GMT
View Forum Message <> Reply to Message

I want to run logistic regression model with Immunization status as the outcome variable with various risk factors. I am using the DHS data 2012-13 from Pakistan. Just wanted to ask if there is a need to account for cluster variation like running GENMOD in SAS? And if it has to be done, what is alternate test for that in SPSS? Thanks

---

Subject: Re: Accounting for cluster variation
Posted by Sarah B on Wed, 14 Oct 2015 16:19:47 GMT
View Forum Message <> Reply to Message

I'm not familiar with GENMOD in SPSS, but I think that what you want to do is adjust for the fact that DHS data do not come from a simple random sample -- they come from a 2-stage stratified, clustered sample design.

DHS has a great video tutorial that explains how to incorporate sample design into your analysis in both SPSS and SAS: see https://www.youtube.com/watch?v=NNg8HD_lKow. Note that this is the fourth video in a series; you probably will want to watch at least some of the earlier videos in this series as well.

Briefly, in SPSS you will need to use the COMPLEX SAMPLES package in SPSS to account for the stratified, clustered sample design. Here's a writeup on how to do this from another thread, which is also a useful read: http://userforum.measuredhs.com/index.php?t=rview&goto=8 50

Trevor-DHS wrote on Wed, 20 March 2013 20:17You can use the Complex Samples procedures in SPSS to achieve the same as using svy in Stata. You first need to set up a Complex Sampling Plan using the CSPLAN command (I recommend creating this using the dropdown menu under Analyze, Complex Samples, Prepare for Analysis, and then pasting it into your SPSS syntax. The parameters you typically need are:
Strata: V023 - or alternatively create your strata variable from a combination of V024 and V025.
Clusters: V021 - typically this is the same as V001, but for a few surveys the Primary Sampling Unit (PSU) is different from the final cluster, and the PSU should be used.
Analysis weight: V005 - don't divide by 1000000 as SPSS expects the weight used with Complex Samples to be an integer. Your "population" size will be a million times too big in your results, but just remember to divide it by 1000000 after your analysis. If you use the weight divided by 1000000, SPSS either rounds or truncates your weight to an integer and your analysis will be wrong.
Estimator type: WR (with replacement) - DHS doesn't use replacement sampling, but to match the DHS results this option is needed.

Once you have created your Complex Samples Plan you can then use one of the Complex Samples Procedures for your analysis. I suggest using the CSDESCRIPTIVES first and reproducing the sampling errors shown in the DHS report for one indicator to ensure that you have the CSPLAN set up properly before you try using one of the other CS procedures such as CSLOGISTIC. [Note that DHS uses confidence intervals of +/-2 SEs, whereas SPSS will use

+/-1.96 SE for the confidence intervals].

Hope that helps!