Subject: Re: Sample weights for blood pressure data
Posted by Bridgette-DHS on Fri, 14 Aug 2020 15:25:17 GMT
View Forum Message <> Reply to Message

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

I was thinking that the surveys would be analyzed separately, even if, for convenience, you appended them all into one file. Now I see why you were asking about denormalization.

I don't like to combine several surveys into a single regression, without at least separate intercepts for the separate surveys. If you believe (and this is a testable hypothesis) that the survey ID is irrelevant, then this implies that each survey is a sample from some mega population and you are estimating the parameters in that mega population. In that case, you could just use v005 as it is coded in each survey.

Another approach would be to weight each sample equally. If you have, say, 6 surveys, then you calculate the total weight in all surveys combined, divide that by 6, and then multiply v005 for each survey by a factor to scale v005 up if that survey has less than 1/6th of the total weight or scale it down if it has more than 1/6th of the total weight. But if you do that, I really have no idea what you are estimating. The surveys are probably from different years and different regions of the world.

It would also be possible to weight each sample in proportion to the population of the country at the time of the survey. If you do this, and you include, say, India or Nigeria or Indonesia, the other countries will have a negligible effect on the results.

I suppose this is a judgment call, but I would vote for keeping the original weights and including (at least) a covariate for survey in your model.