Subject: Sample weights for blood pressure data
Posted by Lukem on Fri, 14 Aug 2020 08:58:23 GMT
View Forum Message <> Reply to Message

Dear DHT Team (and users),

We are planning to investigate the prevalence of hypertension in countries where DHS surveys have collected quantitative or qualitative information on blood pressure. I have several questions with regard to the use of sample weights in these analyses:

1. For countries where only a subsample of women were eligible for blood pressure measurement (usually a subsample living in households eligible for the male survey - examples are Bangladesh 2011 and Nepal 2016), would it be acceptable to simply use individual sample weights (variable v005) or would there need to be some sort of correction for the fact that only a subset of women were eligible for this component of the survey? If so, how would this be accomplished?

2. In the same surveys, where only a subsample of women were eligible for blood pressure measurement, is there a straightforward means to identify women who were eligible for but did not undergo blood pressure testing (in surveys where there is not a variable coding for consent to take blood pressure - which could be used for this purpose)?

3. For regression analyses using multiple appended surveys (looking at socioeconomic variables), having denormalized the sample weights using country-specific demographic information, how would sample weights be applied in the models?

4. Are there any 'special considerations' with regard to the use of clusters and PSUs that would need to be taken into account in these regression analyses?

I appreciate that I have asked several questions here and I would be very grateful for your responses and insights.

Subject: Re: Sample weights for blood pressure data
Posted by Bridgette-DHS on Fri, 14 Aug 2020 14:01:43 GMT
View Forum Message <> Reply to Message

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

We would recommend using v005 as the weight, even for a subsample. I don't know what software you are using, but we prefer Stata. In Stata, with the pweight option, the weights are automatically normalized so that the mean weight for the cases in your model is 1. This should take care of the subsampling issue. If there is additional nonresponse for the subsample that is not random, that will not be taken into account, but I still would use v005.

In Stata, "Not applicable" is indicated by a dot or period (."). Everyone who is not eligible for a question or variable is coded with a dot. If someone was eligible but refused, could not be

located, etc., they should have a numeric code such as 999. You need to be sure to omit those codes when calculating means, etc.

You do not need to "denormalize" the weights! What I said above applies to any model. In Stata, use pweight=v005.

For any estimation model or the calculation of confidence intervals or test statistics you need to use svyset and svy. Please check the forum for references to them.

Subject: Re: Sample weights for blood pressure data
Posted by Lukem on Fri, 14 Aug 2020 14:35:24 GMT
View Forum Message <> Reply to Message

Dear Tom (and Bridgette),

Thank you very much for your rapid response and very useful guidance. We will be using STATA for the analysis and will follow your advice accordingly.

To confirm: Even if combining multiple survey countries within a single regression model, we do not need to make any adjustments to the weight variable v005 for each included country?

Thank you again for your advice.

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.

Subject: Re: Sample weights for blood pressure data
Posted by Lukem on Fri, 14 Aug 2020 16:07:50 GMT
View Forum Message <> Reply to Message

Dear Tom,

Thank you very much for your further advice, this is very useful. We will consider your comments carefully when designing our analysis. I take your point about the overwhelming contribution of India (in the case of blood pressure data) relative to the other studies if weighting each sample in proportion to the country's population.

Many thanks.
