Subject: why weights cannot be used to test for relationship like regression
Posted by mmakali1 on Wed, 02 Mar 2016 22:41:41 GMT

I would like to seek guidance concerning the kenya 2008/2009 DHS guide on sampling weights.What is the rationale on why sampling weights should not be used when performing a relationship like logistic regression ?Am using a merged dataset from household and hiv datasets.Am running a bivariate logistic regression and will also test for effect modification between hiv status and improved water sources and improved sanitation.Attached is an excerpt from the DHS guide on sampling weights below..

Notes and Considerations
1. The sum of the sample weights only equals the number of cases for the entire sample and not for subgroups such as urban and rural areas.
2. Where there are no differential probabilities, weights may not be calculated since weights based just on response rates usually make little difference in results.
3. Use of sample weights is appropriate when representative levels of statistics are desired, such as percentages, means, and medians.
4. Use of sample weights is inappropriate for estimating relationships, such as regression and correlation coefficients.
5. Use of sample weights biases estimates of confidence intervals in most statistical packages since the number of weighted cases is taken to produce the confidence interval instead of the true number of observations. For oversampled areas or groups, use of the sample weights will drastically overestimate sampling variances and confidence intervals for those groups.

Subject: Re: why weights cannot be used to test for relationship like regression
Posted by Bridgette-DHS on Thu, 10 Mar 2016 14:14:41 GMT

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I believe you have been looking at the Guide to DHS Statistics, which advises against the use of weights in statistical models. I apologize for the confusion. This is an outdated recommendation.

The current advice is that weights should be used for models, calculation of rates, percentages, means, etc.--that is, for everything. The Guide to DHS Statistics is being gradually updated, and that old advice is being changed.

I think there are two reasons why that recommendation got into the Guide in the first place. One is that many economists are opposed to the use of weights. They have their reasons, and I respect their reasons, but do not believe they apply to DHS data, most of which is used for descriptive purposes. The second reason is that until a few years ago (I would say less than 20 years ago) computer software could not be relied upon to use weights correctly in the calculation of standard errors. There was a time when SPSS and Stata, for example, gave different results when you would have expected the same results. Fortunately, such discrepancies have been eliminated. Stata, for example, can be relied on to give the correct robust standard errors, adjusted for strata and clusters as well as weights.